1962

Application of certain statistical techniques to the analysis of core samples.

Van Scoyoc, James S.

University of Kansas

Application Of Certain Statistical Techniques To The Analysis Of Core Samples

[illegible]MES [illegible] [illegible]COY[illegible]

[illegible]

APPLICATION OF CERTAIN STATISTICAL TECHNIQUES TO THE ANALYSIS OF CORE SAMPLES

by

James S. Van Scoyoc
B.S., United States Naval Academy, 1953

Submitted to the Department of Chemical and Petroleum Engineering and the Faculty of the Graduate School of the University of Kansas in Partial Fulfillment of the Requirements for the Degree of Master of Science.

PREFACE

The author has attempted to illustrate in a somewhat brief manner the application of certain statistical techniques to the analysis of core sampling data. The statistical areas of frequency distributions, analysis of variances, and to a lesser degree, sampling, provide the basis for the study.

A convenient reference system is used in the thesis. All equations are numbered as (Chapter, Number). Thus, equation (2-4) signifies the fourth equation of the second chapter. The notes are listed together before the bibliography and are numbered consecutively within each chapter. All mathematical notations and significant terms used are listed and defined in Appendix A for ready reference.

Actual field data was obtained for the study. The author is indebted to the following people who provided the field data upon which the project was based: Mr. Mack C. Colt and Mr. Wendell Weatherby of Iola, Kansas; Mr. Schermerhorn of Tulsa, Oklahoma; Mr. Ray Plummer of Chanute, Kansas; and Mr. Carl Pate and the Oil Field Research Laboratory of Chanute, Kansas.

The author wishes to express his appreciation to Dr. Charles F. Weinaug for his overall direction of the graduate program which led to the thesis, and to the Bureau of Supplies and Accounts, United States Navy, whose sponsorship made the thesis possible.

The author is most deeply indebted to Dr. Floyd Preston whose close personal guidance and wise counsel were instrumental in bringing the thesis to completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure		Page

CHAPTER I
INTRODUCTION

Background

Statistics may be considered in two senses. One conception of statistics is that of a collection of numerical or quantitative data, i.e., figure data, such as numerical data on births or unemployment. Statistics in the second sense is less well known, and refers to the techniques of analyzing data for decision-making. It may be thought of as the science of decision-making in the face of uncertainty.

The application of statistics in this second sense to petroleum engineering problems is the purpose of this thesis. The petroleum engineer by the very nature of the realm in which he must operate is continually faced with making decisions based upon fragmentary and inconclusive information concerning the sub-surface of the earth. Because of the extreme complexity of the geological processes of erosion, material transport, deposition and burial of material, the porous media forming petroleum reservoirs are extremely heterogeneous. Physical properties can vary extensively from place to place within individual reservoirs. The extreme cost of sampling these reservoirs forces the engineer to construct mental and mathematical models from fragmentary information. Mathematical statistics would seem to provide a valuable tool for such model construction. The utilization of the science of statistics to provide systematic analysis

techniques to data that is available may provide additional relevant information as an assist to the decision-making process.

It is the purpose of this thesis to test the applicability of certain statistical techniques for creating mathematical models of property variation within petroleum reservoirs and to show how certain statistical techniques can be used to extend the interpretation of fragmentary data. The study is limited to the analysis of core samples. Statistical techniques are applied to actual field data obtained from five fields. Specific results are presented to demonstrate calculational techniques.

The first approach in the study is to examine the probable frequency distributions of the various properties, testing to see if property distributions within individual fields satisfy the Gaussian normal distribution and if not, to describe the distributions by a family of generalized frequency curves known as the Pearson system of frequency curves. An initial study of the distributions is considered essential, prior to the application of the other statistical techniques such as sampling and analysis of variances, since these depend upon the nature of the frequency distribution of the data under study.

Origin of Statistics

The origin of the modern science of statistics may be traced to the mid-seventeenth century when two astute French mathematicians, Pascal and Fermot, were presented with a

problem involving a game of chance and the interpretation of the probabilities associated with it. Their work led to solutions, not only of the problems proposed, but of more general ones. The methods employed by Pascal may be said to represent the beginning of the mathematics of probability, about which modern statistical theory centers today. The publication by Laplace in 1812 of the epoch-making "Theorie Analytigue des Probabilities" laid a firm foundation for this theory.[1]

Statisticians in general from Pascal onward sought a method of describing the nature of the distributions of chance effects. Table I[2] shows the most notable of the methods developed to portray chance effects from the time of Pascal until Karl Pearson set forth his system of generalized frequency curves in 1895.

The usual purpose of frequency distributions is to represent a sample of actual data drawn from a much larger or even infinite population. Even though a sample may be composed of a relatively small finite number of observations, it may be reasonably representative of the larger universe from which it was drawn. Since it is virtually impossible to measure all the items comprising a universe, it is necessary to form a notion of the larger group from the study of a sample. Thus by constructing a hypothetical infinite population of which the actual data is regarded as constituting a random sample, an understanding of the law of distribution of chance effects of the hypothetical

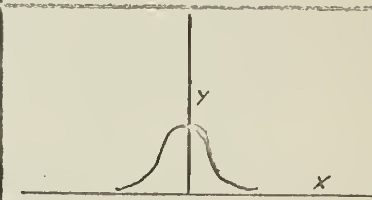
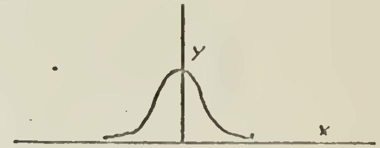
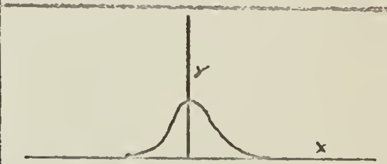
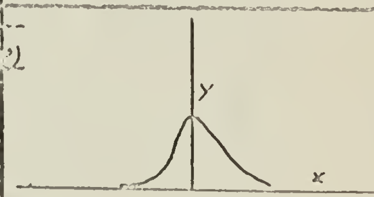
Date	Name	Nature of Distribution of Chance Effects	Expression of the Law of Chance: Analytical Form	Expression of the Law of Chance: Graphical Form
1756 1773	Simpson and LaGrange	Symmetry in Distribution. Limited Range of Variation.	$y = \pm mx + c$	
1774	Laplace	Symmetry in Distribution. Unlimited Range of Variation.	$y = \frac{m}{2} e^{\pm mx}$	
1778	Daniel Bernoulli	Symmetry in Distribution. Limited Range of Variation.	$y = \sqrt{r^2 - x^2}$	
1809	Gauss	Symmetry in Distribution. Unlimited Range of Variation. Observed chance effects distributions in all experimental data are governed by the law of chance expressed in Gaussian Normal Law of Errors.	$y = \frac{h}{\sqrt{2\pi}} e^{-h^2x^2}$	
1812	Laplace	Symmetry in Distribution. Unlimited Range of Variation. Observed chance effects distributions represent an approximation to the Law of Chance.	$y = \frac{1}{\sqrt{2\pi P_2}} e^{-\frac{x^2}{2P_2}}$	
1837 1832-1849	Poisson and Quetelet	An apparent casual irregularity in observed events is governed by the universal law of large numbers. Symmetrical as well as assymetrical distribution of final effects of chance.	$y = \frac{1}{\sqrt{2\pi P_2}} e^{-\frac{x^2}{2P_2}} = \varphi(x)$ $y = \frac{e^{-m} m^x}{x!} = \psi(x)$	
1879 1889 1905	Gram Thiele Charlier	(Symmetrical Distributions. (Assymetrical Distributions. (Limited Range (Unlimited Range	$y = a_0\varphi^0(x) + a_1\varphi'(x) + a_2\varphi''(x) + \dots$ $\varphi^0 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2\sigma^2}}; \varphi^n = \frac{d^n\varphi}{dx^n}$ $y = c_0\psi(x) + c_1\Delta'\psi(x) + \dots$ $\Delta'\psi(x) = \psi(x) - \psi(x-1)$	
1895 1901 1916	Pearson	Assymetry in Distribution. Finite Range of Variation. Non-existence of one Universal Law of Chance.	$\frac{1}{y}\frac{dy}{dx} = \frac{a_0 + x}{c_0 + c_1 x + c_2 x^2}$	Too numerous to give here. Refer to Chapter II.

TABLE 1 - How Chance Effects Have Been Portrayed.

population may be obtained and specified by a few parameters.

For example, to gain an understanding of the porosity characteristics of an area, it may be considered that the earth contains an infinite number of porosity measurements where it is impossible to measure each and every one of the porosities. By fitting a curve to a frequency distribution of the actual data obtained from a core sample, it is attempted to describe what appears to be the general form of the curve for the entire porosity population.

For statistical work the normal or Gaussian curve is probably the best known and most heavily relied upon frequency distribution of those shown in Table I, particularly in the theory of sampling. However, this distribution function does not apply to skewed frequency distributions, although numerous sets of data defy the normal curve and exhibit markedly skewed distributions. The Gaussian school of statisticians regarded skewness as a by-product of sampling and believed that skewness could be made to disappear completely if an infinite number of observations were available.

With the recognition that the normal curve was not sufficient to characterize all natural observations, it became apparent that it was necessary either to devise methods of describing the most conspicuous departures from the normal distribution or to devise generalized frequency curves to describe distributions as they actually exist in the observational sphere.

Karl Pearson followed the latter course and showed that

a set of frequency curves could be obtained by assigning values to the parameters in a certain first order differential equation which has its basis in the theory of probability. This approach is covered in Chapter II and the application of these curves to the actual field data is shown in Chapter III.

Core Samples

The taking of core samples from a reservoir has been an accepted practice for the past hundred years. At first, well samples and cores had but one purpose--to locate oil.[3] Thus it was necessary to take a sample from every well drilled. Today, however, with the advent of other techniques such as electric logging to aid in locating the oil the primary reason for taking samples has shifted to serve as a source of information about the reservoir and its contents. At present, core samples provide certain numerical parameters by which the field may be described. The most common is the arithmetic or weighted average of the various properties. The range and variances of the distribution of the properties are other easily obtained and useful parameters describing a field.

A simple histogram showing the distribution presents visually the characteristics of a field. A cumulative frequency curve on graph paper will present the distribution and allow easy determination of such parameters as the median and possibly the mode. From a cumulative frequency curve an estimate of the percentage of a distribution which is above

a specified minimum point may be obtained. The range of a variable within any set quartile is likewise easily determined. In summary, graphical methods of statistical data presentation permit certain numerical parameters to be obtained with relative ease.

Since it is no longer necessary to take samples from every well, it is desired to determine the number of wells that should be core-sampled to provide information needed with an acceptable probability of obtaining reliable results. At present the number of core samples to be taken is determined somewhat intuitively with a wide variation of opinion as to what is the necessary number. Are there statistical techniques available to serve as a guideline in determining how much information is needed and how such data should be interpreted?

The method of interpretation of core data may be paramount since raw data by itself, irregardless of the amount available, often supplies much information that is irrelevant and immaterial. It is the object of statistical processes employed to exclude this irrelevant information and to isolate the whole of the relevant information contained in the data.

Data and Techniques

The science of statistics consists of (1) collecting, (2) presenting, and (3) analyzing quantitative data. The data used for this study consist of the physical properties

of oil fields, namely permeability (K), porosity (ϕ), oil saturation (S_o), and water saturation (S_w), as obtained from core samples taken from five different fields. Table II[4] shows the type and amount of data considered. The five fields will be referred to as Field 1, Field 2, Field 3, Field 4, and Field 5. The numbering system has no significance other than the fields being numbered in sequence as data were obtained for this study.

Fields 1, 2, 4, and 5 are located in southeast Kansas while Field 3 is in northeast Oklahoma. Field 1 is in County "A" while Fields 2, 4, and 5 are in County "B".

The core samples for all five fields were taken and analyzed by the same laboratory with the same coring and analyzing techniques used for all fields. Thus even though there may have been errors made in arriving at the absolute values of measurements of the different properties, especially with regard to the fluid saturations, the errors may be considered consistent, and a relative comparison of the data may be made with a certain degree of confidence.

TABLE II Summary of Core Sampling Data

Field	No. of Wells	No. of K Samples	No. of ϕ Samples	No. of S_o Samples	No. of S_w Samples
1	40	1303	794	794	794
2	14	630	345	345	345
3	7	195	132	132	132
4	19	-	316	316	201
5	101	-	1673	1672	1129

Standard coring techniques were employed in taking the vertical cores of approximately 20 feet in length. Measurements of each property were made approximately every six inches in the oil productive section with approximately twenty samples being obtained for each well. The depth to the pay zone for the different fields varied somewhat, but they all could be considered as shallow fields with pay zones at depths between six hundred and a thousand feet.

The well spacing in most instances was approximately four hundred feet. In each field more than 50 percent of the wells were cored, with coring data available for all properties with the exception of permeabilities for Fields 4 and 5.

A map of each of the first three fields showing the locations of the wells cored is given in Appendix C.

The above mentioned data were used for the following statistical studies which constitute both the method of and justification for this thesis:

1. Analytical Fitting of Data to the Pearson Generalized Frequency Curves. The study includes the technique for selecting the appropriate Pearson type curve, for fitting the data to the selected curve and for measuring the goodness of fit of the data to the curve. In addition the data were fitted to the normal or Gaussian Curve and its goodness of fit determined. Permeability data were converted to logrithms and the resultant distributions were analyzed by the above methods.

2. An Analysis of Variance Study of Well Property Means and Variances.
3. Use of Certain Sampling Methods as a Way of Estimating Certain Population Parameters from Point Estimates.
4. The Application of Additional Statistical Techniques to Core Analysis Are Briefly Discussed.

Statistics, in its many ramifications, is an exceedingly complex subject and much too involved to be thoroughly covered in a paper of this nature. The techniques presented are by no means the only methods of analysis available. In order to permit the reader without a thorough background in theoretical statistics to gain an understanding of the material presented, the discussion of the theoretical proofs and principles involved have been held to a minimum.

CHAPTER II
FREQUENCY CURVES

Introduction

One of the most important practical problems in mathematical statistics is the obtaining of a relatively simple yet accurate representation of the frequency distribution of any set of data under consideration. Some knowledge of the frequency distribution of a set of data should be obtained before a statistical analysis is attempted.

There are essentially three methods of describing frequency distributions of one variable; namely, the graphical method, the method of averages and dispersions, and the method of theoretical frequency functions or curves. These three methods of describing frequency curves will be briefly compared to their relative merits.

The graphical method allows a large amount of data to be condensed to an easily presentable form. An inherent weakness of this method is the inability to quantitatively compare distributions of two or more sets of data. One may state that two distributions are somewhat the same or that they are somewhat different, but the degree of sameness or difference cannot be quantified by observation alone. This lack of numerical description of the distribution by the graphical method precludes its use as a comparing method except in the most elementary studies.

Figure 1 shows two histograms, one representing the frequency distribution of the porosity of Field 1 and the

other that porosity of Field 2. From the first method, that of graphical presentation, one might conclude that the two histograms were somewhat similar. They both are slightly skewed in a negative direction with approximately the same distribution. One might assume then that the distribution pattern for porosity for the two fields are nearly the same. Such an assumption though would be misleading, for through the method of theoretical frequency distributions used herein, it will be shown that these two distributions are more different than a simple visual comparison would indicate.

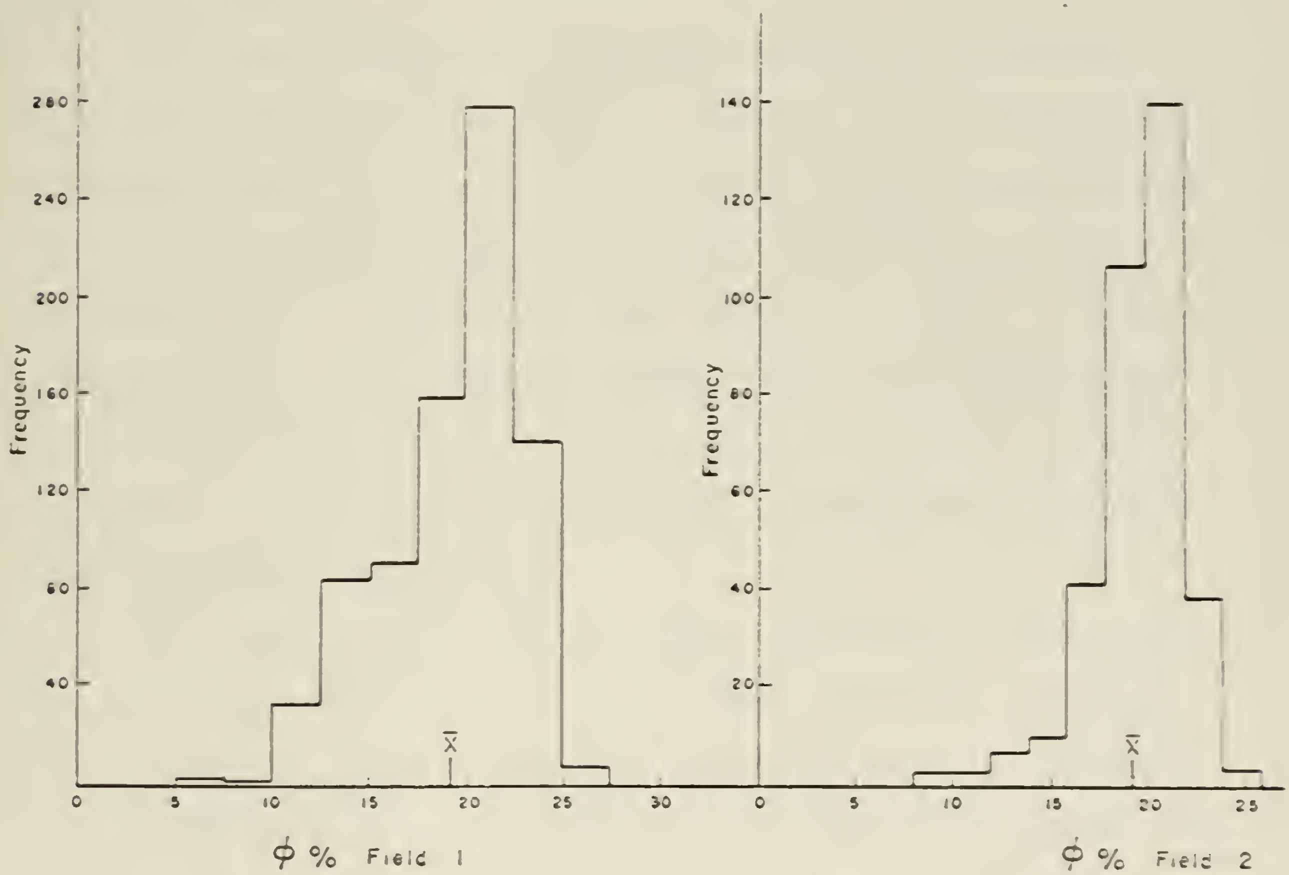


FIGURE 1 Histogram of Porosity Distributions--Fields 1 and 2

The second method, involving the use of averages and dispersion, does give a numerical description of the data in Figure 1, but it does not give a functional relation between the values of the variable X and the corresponding frequencies. The numerical description in terms of $\bar{X}$ and σ where:

$\bar{X}$ = Arithmetic mean

σ = Standard deviation

Field	$\bar{X}$	σ
1	19.4%	1.46
2	19.7%	1.17

This numerical description would further indicate that the distributions are nearly the same which again would be somewhat misleading.

The third method, an analytical method of describing frequency distributions shows that Field 1 is a generalized frequency distribution of the Pearson Type I_B, whereas Field 2 is of the Pearson Type IV_B and may be described by the parameters $\alpha_3^2 = .725$, $\delta = -.294$, and $\alpha_3^2 = 1.86$, $\delta = .141$, respectively. This method indicates that the distributions of the two fields are not similar which is contrary to what one would assume by use of the first two methods of describing frequency distributions.

The reader should not be unduly concerned at this point as to how the values of α_3^2 and δ were determined and what Type I_B and IV_B mean as the following sections will present a thorough discussion of Pearson's generalized frequency curves.

Pearson's Generalized Frequency Curves

After it was recognized that the Gaussian or normal

curve failed to describe the distribution of many of the observed data, Pearson proceeded to develop generalized frequency curves that could characterize the various types of unimodal frequency distributions encountered.

In deciding on a system of curves for describing frequency distributions, it was realized that:

1. Any expression used should be a graduation formula, i.e., it must remove the roughness of the data.
2. An expression should not involve too many high moments to calculate constants, for thereby accuracy is reduced.
3. There should be a systematic method of analysis applicable to all possible types of frequency distributions.

Then, considering the most obvious characteristics of frequency distributions, it may be considered that they generally start at zero, rise to maximum, and then fall at the same or often at a different rate. At the end of the distribution there is often high contact. Mathematically, the above implies that a series of equations $Y = f(x)$ or $Y = f(t)$, must be chosen so that in each equation of the series dy/dx or $dy/dt = 0$ for certain values of x or t, namely at the maximum and at the end of the curve where there is contact with the axis of x or t.

The above suggests that the frequency function may be represented as a solution of the differential equation:

$$\frac{dy}{dt} = \frac{Y(a-t)}{f(t)} \tag{2-1}$$

since:

a. For a value of t, $t = a$, $dy/dt = 0$ and the required

maximum is obtained and,

b. As Y approaches zero, the derivative dy/dt also approaches zero thus giving contact at one end of the curve.

Assuming that f(t) may be expanded in a converging power series, equation (2-1) may be written:[1]

$$\frac{1}{Y}\frac{dy}{dt} = \frac{a - t}{b_o + b_1 t + b_2 t^2 + \ldots\ldots} \tag{2-2}$$

where the mean of the distribution is taken as the origin, and the abscissae are measured in units of the standard deviation such that:

$$t = \frac{\text{distance from origin}}{\sigma} \tag{2-3}$$

The above suggests that significant frequency functions Y = f(t) may be found among the solutions of (2-2) subject to the following restrictions:[2]

$$\alpha_o = \int_{b_2}^{b_1} f(t)\, dt = 1 \tag{2-4}$$

$$\alpha_1 = \int_{b_2}^{b_1} t\, f(t)\, dt = 0 \tag{2-5}$$

$$\alpha_2 = \int_{b_2}^{b_1} t^2 f(t)\, dt = 1 \tag{2-6}$$

where:

$$\alpha_n = \frac{\mu_r}{\sigma^n} = \int_{-\infty}^{\infty} t^n f(t)\, dt \tag{2-7}$$

and,

(2-7a) $$\alpha_3 = \frac{\mu_3}{\sqrt{\mu_2^{\,3}}} \qquad \alpha_4 = \frac{\mu_4}{\mu_2^{\,2}}$$

See Chapter III for definition of the moments μ_2, μ_3, ... μ_n. Clearing (2-2) of fractions, multiplying through by t^n and integrating over the range r to s (where r and s are the extremes of the range of variation for t) with respect to t gives:[3]

(2-8) $$\left[a\int t^n Y dt - b_o\int t^n dy - b_1\int t^{n+1} dy - b_2\int t^{n+2} dy - \int t^{n+1} Y dt\right]_{t=r}^{S} = 0$$

But through integration by parts:

(2-8a) $$\left[\int t^n\, dy\right]_{t=r}^{S} = \left[t^n\, Y - n\int t^{n-1}\, Y\, dt\right]_{t=r}^{S}$$

and if the frequency function, when multiplied by t^n vanishes at the limits of the distribution, r and s, we have that:[4]

(2-9) $$\int_{t=r}^{S} t^n\, dy = -n\, N\, \alpha_{n-1}$$

which leads to the recursion formula for moments:[5]

(2-9a) $$\alpha_n\, a + n\, \alpha_{n-1}\, b_o + (n+1)\, \alpha_n b_1 + (n+2)\, \alpha_{n+1}\, b_2 = \alpha_{n+1}$$

Giving n successively the values 0, 1, 2 ..., from (2-8) and (2-9) noting that N the total frequency cancels out and letting $\alpha_o = 1$, $\alpha_1 = 0$, $\alpha_2 = 1$, one obtains:[6]

$$
\begin{array}{llllll}
 & a + & & b_1 & & + \ldots = 0 \\
 & & b_o + & & 3\, b_2 & + \ldots = 1 \\
(2\text{-}10) & a + & & 3\, b_1 & + 4\, \alpha_3\, b_2 & + \ldots = \alpha_3 \\
 & \alpha_3\, a + & 3b_o + 4\alpha_3 b_1 & & + 5\, \alpha_4\, b_2 & + \ldots = \alpha_4 \\
 & \alpha_n\, a + n\, \alpha_{n-1}\, b_o + (n+1)\, \alpha_n b_1 & & & + (n+2)\, \alpha_{n+1}\, b_2 & + \ldots = \alpha_{n+1}
\end{array}
$$

Assuming that $f(t)$ converges so rapidly that terms involving the third and higher powers of t may be neglected, a simultaneous solution of (2-10) yields:

$$
(2\text{-}11) \qquad a = \frac{-\alpha_3}{2(1+2\delta)}, \qquad b_1 = \frac{\alpha_3}{2(1+2\delta)}
$$

$$
b_o = \frac{2+\delta}{2(1+2\delta)}, \qquad b_2 = \frac{\delta}{2(1+2\delta)}
$$

where,[7]

$$
(2\text{-}12) \qquad \delta = \frac{2\,\alpha_4 - 3\,\alpha_3^{\,2} - 6}{\alpha_4 + 3}
$$

The value of a, where:

$$
(2\text{-}13) \qquad a = \frac{-\alpha_3}{2(1+2\delta)}
$$

represents the distance between the mean and the mode, which is defined by Pearson as the skewness of the distribution.

Thus from the differential equation:

$$
(2\text{-}2) \qquad \frac{dy}{Y\,dt} = \frac{a-t}{f(t)} = \frac{a-t}{b_o + b_1\, t + b_2\, t^2}
$$

which has its basis in the theory of probability, the parameters a, b_o, b_1, and b_2 have been determined in terms of the moments and expressed in terms of α_3 and δ. This

transformation is desirable to permit the use of the (α_3, δ) chart for identification of the appropriate type curve that fits a set of data. With the above parameters defined, Pearson's family of generalized frequency curves may be developed.

In the family of curves there are three main types, nine transitional types which are special conditions of the three main types and the normal curve.

Integration of the Differential Equation

For $\delta \neq 0$, $b_2 \neq 0$, the denominator $b_o + b_1 t + b_2 t^2$ is a quadratic which can be written in the form $b_2(t - r_1)(t - r_2)$.

Thus,

(2-14) $$\frac{1}{Y}\frac{dy}{dt} = \frac{a - t}{b_o + b_1 t + b_2 t^2} = \frac{a - t}{b_2(t - r_1)(t - r_2)}$$

where:

(2-15) $$r_1 = \frac{-b_1 + \sqrt{b_1^2 - 4b_o b_2}}{2 b_2} \text{ and } r_2 = \frac{-b_1 - \sqrt{b_1^2 - 4b_2 b_o}}{2 b_2}$$

Upon substitution of α_3 and δ in (2-15), for a, b_o, b_1, and b_2:

(2-15a) $$r_1 = -\frac{\alpha_3 + \sqrt{\alpha_3^2 - 4\delta(\delta+2)}}{2\delta} = \frac{-\alpha_3 + \sqrt{D}}{2\delta}$$

where:

$$D = \alpha_3^2 - 4\delta(\delta+2) \text{ and}$$

(2-15b)

$$r_2 = -\alpha_3 \frac{-\sqrt{\alpha_3^2 - 4\delta(\delta+2)}}{2\delta} = \frac{-\alpha_3 - \sqrt{D}}{2\delta}$$

Also, by the method of partial fractions,[8]

$$\text{(2-16)} \qquad \frac{a - t}{b_2(t-r_1)(t-r_2)} = \frac{1}{b_2}\left[\frac{A_1}{t-r_1} + \frac{B}{t-r_2}\right] \text{ where}$$

$$\text{(2-17)} \qquad A_1 = \frac{a-r_1}{r_1-r_2} \text{ and } B = \frac{a-r_2}{r_2-r_1}$$

Hence from (2-14):

$$\text{(2-18)} \qquad \frac{1}{Y}\,dy = \frac{1}{b_2}\frac{(a-r_1)}{(r_1-r_2)}\frac{dt}{(t-r_1)} + \frac{1}{b_2}\frac{(a-r_2)}{(r_2-r_1)}\frac{dt}{(t-r_2)}$$

Upon integration of (2-18):

$$\text{(2-19)} \qquad \text{Log } Y = \frac{1}{b_2}\left(\frac{a-r_1}{r_1-r_2}\right)\log(t-r_1) + \frac{1}{b_2}\left(\frac{a-r_2}{r_2-r_1}\right)\log(t-r_2)+\log C$$

Hence:

$$\text{(2-20)} \qquad Y = C(t-r_1)^{\frac{1}{b_2}\left(\frac{a-r_1}{r_1-r_2}\right)}(t-r_2)^{\frac{1}{b_2}\left(\frac{a-r_2}{r_2-r_1}\right)}$$

$$\text{(2-21)} \qquad \text{Let } m_1 = \frac{1}{b_2}\left(\frac{a-r_2}{r_1-r_2}\right) \text{ and } m_2 = \frac{1}{b_2}\left(\frac{a-r_2}{r_2-r_1}\right)$$

Thus (2-20) reduces to:

$$\text{(2-22)} \qquad Y = C(t-r_1)^{m_1}(t-r_2)^{m_2}$$

Upon substitution of α_3 and δ in (2-21)

$$\text{(2-23)} \qquad \begin{aligned} m_1 &= \left(\frac{1+\delta}{\delta}\right)\left(\frac{\alpha_3}{\sqrt{D}}\right) - \left(\frac{1+2\delta}{\delta}\right) \text{ and} \\ m_2 &= -\left(\frac{1+\delta}{\delta}\right)\left(\frac{\alpha_3}{\sqrt{D}}\right) - \left(\frac{1+2\delta}{\delta}\right) \end{aligned}$$

For $-4 < \delta < 0$, the r's are real and opposite in sign; for $\delta > 0$, and $\alpha_3^{\,2} < 4\delta(\delta+2)$, the r's are complex; and for $\delta > 0$

and $4\delta(\delta+2) < \alpha_3^2$ the r's are real and of the same sign. These three conditions with the additional condition that $\alpha_3 \neq 0$ establish the criteria for the three main types of Pearson's frequency functions designated Type I, IV, and VI.[9] The boundaries of these areas, the curve $(2+3\delta)\ \alpha_3^2 = 4(1+2\delta)^2\ (2+\delta)$ which intersects the Type I and Type VI areas and the line $\delta = -1/2$ contains the points which correspond to the transitional types. The numbering system of the main types, i.e., I, IV, VI, is that established by Pearson and is used in this study to provide a standard reference to other literature on the family of curves.

Analysis of Data

For the purposes of analyzing the field data under consideration, it will be shown that Pearson's three main types of curves: Type I, Type IV, Type VI, plus the transitional Type III and the normal curve will suffice to describe the data. Therefore, the development of the other transitional types will not be described in detail. Only the equations and conditions are shown in Table III.[10] On the following pages derivations are given for the five types of frequency distribution functions used in this thesis.

Type I

When the r's in (2-14) are opposite in sign, (2-22) is written as:

(2-24) $$Y = C(t-r_1)^{m_1}\ \ (r_2-t)^{m_2}$$

Equation (2-24) is called Type I of the Pearson system. The conditions on α_3 and δ are:

$$\alpha_3 \neq 0, \; -1 < \delta < 0 \; (\delta \neq -1/2), \; (2+3\delta)\,\alpha_3^2 \neq 4(1+2\delta)^2\,(2+\delta)$$

C is determined by setting the area i.e., the total probability equal to unity or,

$$C \int_{r_1}^{r_2} (t-r_1)^{m_1} \, (r_2-t)^{m_2} \, dt = 1 \tag{2-25}$$

Substituting:

$$W = \frac{t-r_1}{r_2-r_1} \tag{2-26}$$

$$C \int_0^1 W^{m_1} \, (r_2-r_1)^{m_1 + m_2 + 1} \, (1-W)^{m_2} \, dW = 1 \tag{2-27}$$

Hence from the definition of the Beta function:[11]

$$C(r_2-r_1)^{m_1 + m_2 + 1} \, \beta(m_1 + 1, \, m_2 + 1) = 1 \tag{2-28}$$

Thus C may be expressed either in terms of the Beta function, or alternately in terms of the Gamma function, through use of an identity given by Whitaker and Watson:[12]

$$C^{13} = \frac{1}{(r_2-r_1)^{m_1+m_2+1} \, \beta(m_1+1, M_2+1)} = \tag{2-29}$$

$$\frac{1 \; \Gamma(m_1+m_2+2)}{(r_2-r_1)^{m_1+m_2+1} \, \Gamma(m_1+1)\Gamma(m_2+1)}$$

The other parameters are:

$$r_1 = \frac{-\alpha_3 + \sqrt{D}}{2\ \delta}, \qquad r_2 = \frac{-\alpha_3 - \sqrt{D}}{2\ \delta}$$

$$m_1 = -\left(\frac{1+\delta}{\delta}\right)\left(1 \frac{-\alpha_3}{\sqrt{D}}\right) - 1, \qquad m_2 = -\left(\frac{1+\delta}{\delta}\right)\left(1 + \frac{\alpha_3}{\sqrt{D}}\right) - 1$$

for $\alpha_3 > 0$, $r_1 < 0 < r_2$ and $|r_1| < |r_2|$.

The range of the curve is (r_1, r_2). The curve will be U-shaped if both m's are < 0, J-shaped if the m's are opposite in sign, and bell-shaped if both m's are > 0.[14]

Type IV

Conditions: $\alpha_3 \neq 0$, $\delta > 0$, $\alpha_3^2 < 4(\delta+2)$. The conditions imply that the r's in (2-14) are complex and therefore the second main type of curve, Type IV, may be determined. Thus (2-15) can be written:

$$r_1 = \frac{-\alpha_3 + i\sqrt{-D}}{2\ \delta} = -r + iS \text{ where } r = \frac{\alpha_3}{2\delta} \text{ and } S = \frac{\sqrt{-D}}{2\delta}$$

$$r_2 = \frac{-\alpha_3 - i\sqrt{-D}}{2\ \delta} = -r - iS$$

$$m_1 = \frac{vi}{2} - m \qquad m_2 = \frac{-vi}{2} - m$$

where:

$$v = -2\left(\frac{1+\delta}{\delta}\right)\frac{\alpha_3}{\sqrt{-D}} \text{ and } m = \frac{1+2\delta}{\delta}$$

Thus:

(2-22) $\quad Y = C(t-r_1)^{m_1}\ (t-r_2)^{m_2}$ becomes

(2-30) $$Y^{15} = C[(t+r)^2 + S^2]^{-m} \left(\frac{t+r-iS}{t+r+iS}\right)^{\frac{vi}{2}}$$

and since:[16]

$$\left(\frac{a-bi}{a+bi}\right)^{\frac{ci}{2}} = e^{c \tan^{-1} b/a} = e^{c(\frac{\pi}{2} - \tan^{-1} a/b)}$$

the frequency function can be written:

(2-31) $$Y = C[(t+r)^2 + S^2]^{-m} \quad e^{-v \tan^{-1}(\frac{t+r}{S})} \quad e^{v\frac{\pi}{2}}$$

To determine C on setting the area of the curve over the interval $(-\infty, \infty)$, equal to unity, Craig[17] shows that:

(2-32) $$C^{13} = \frac{S^{2m-1}}{G(2m-2, v)}$$

where:

(2-33) $$G(2m -2, v) = \int_0^{\pi} \sin^{(2m-2)} \phi \, e^{v \phi} \, d\phi$$

The term ϕ is defined as:

(2-33a) $$\phi = \left(\frac{\pi}{2} - \tan^{-1} \frac{t+r}{S}\right)$$

The function $G(2m -2, v)$ is obtainable in tabular form from Pearson.[18]

For this thesis, the function $G(2m -2, v)$ was generated by a special numerical integration procedure involving Gaussian coefficients.[19]

Type VI

Conditions: $\alpha_3 \neq 0$, $\delta > 0$, $\alpha_3{}^2 > 4\delta(\delta+2)$. The conditions imply that the r's in (2-14) are real and of the

same sign, thus the third main type of curve, Type VI, is obtained.

(2-22) $$Y = C(t-r_1)^{m_1} \ (t-r_2)^{m_2}$$

An alternative simplified form is:

(2-34) $$Y = C(Z^{m_2})(Z - \alpha)^{m_1}$$

Craig[17] shows that:

(2-35) $$C^{13} = \frac{\Gamma(-m_2)}{\Gamma(m_1+1)\ \Gamma(-m_2 - m_1 - 1)\ \alpha^{(m_1+m_2+1)}}$$

where:

(2-36) $$Z = t-r_2$$

(2-37) $$\alpha = r_1-r_2$$

Range of curve is (r_1, ∞).

Normal Type Curve

Conditions: $\alpha_3 = \delta = 0$. The original differential equation:

(2-2) $$\frac{1}{Y}\frac{dy}{dt} = \frac{a - t}{b_o + b_1\ t + b_2\ t^2}$$

reduces to:

(2-38) $$\frac{1}{Y}\frac{dy}{dt} = -t$$

when:

$\alpha_3 = \delta = 0$, since from equation (2-11)

$$a,\ b_1,\ b_2 = 0 \text{ and } b_o = 1$$

Hence:

(2-38) $$\frac{dy}{Y} = -t\ dt$$

Upon integration:

(2-39) $$\operatorname{Log} Y = -\frac{t^2}{2} + \log C$$

(2-40) $$Y = C\ e^{\frac{-t^2}{2}}$$

where:

$$C^{20} = \frac{N}{\sqrt{2\pi}\ \sigma}$$

where N = total number of observations.

Type III

Conditions: $\alpha_3 \neq 0$, $\delta = 0$. From equation (2-11) for $\delta = 0$:

$$a = -\frac{\alpha_3}{2} \qquad b_1 = \frac{\alpha_3}{2}$$

$$b_o = \frac{2}{2} = 1 \qquad b_2 = 0$$

Therefore the differential equation:

(2-2) $$\frac{1}{Y}\frac{dy}{dt} = \frac{a - t}{b_o + b_1 t + b_2 t^2}$$

becomes:

(2-41) $$\frac{1}{Y}\frac{dy}{dt} = \frac{(-\frac{\alpha_3}{2} - t)}{1 + \frac{\alpha_3}{2} t} = \frac{-(\frac{\alpha_3}{2} + t)}{1 + \frac{\alpha_3}{2} t}$$

which yields after integration:[21]

(2-42) $$Y = C\left(1 + \frac{\alpha_3}{2} t\right)^{\frac{4}{\alpha_3^2} - 1} e^{-\frac{2}{\alpha_3} t}$$

Let $A = \frac{2}{\alpha_3}$

and:

(2-43) $$Y = C \frac{1}{A}(A+t)^{A^2-1} e^{-At}$$

where:

$$C \frac{1}{A} = C_1$$

and Craig[16] shows that:

(2-44) $$C_1^{13} = \frac{A^{A^2}}{e^{A^2} \, \Gamma A^2}$$

Since it is the purpose of this thesis to utilize Pearson's frequency curves in fitting the field data under consideration, rather than to develop the mathematical theory upon which the curves are based, which is a complete thesis in itself, only limited discussion of the development of the curves has been presented. For a more complete coverage of the development of these curves the following sources may be consulted. The foregoing discussion and derivations represent a synopsis of information given in the following references:

a. Annals of Mathematical Statistics, Volume VII, 1936, "A New Exposition and Chart for the Pearson System of Frequency Curves" by Cecil C. Craig.

b. Frequency Curves and Correlation, by W. P. Elderton, C. Cambridge, 1938.

c. "Karl Pearson's System of Generalized Frequency Curves", by Arnold M. Wedel, Thesis, University of Kansas, 1948.

d. Handbook of Mathematical Statistics, "Frequency Curves", by H. C. Carver, pp. 92-119, 1924.

The entire family of curves is shown in Table III. The main types are shown followed in order by the transitional types associated with a particular main type. This study has been limited to the use of the three main types, the transitional Type III, and the Normal Curve. These five curves adequately describe the field data used for this study.

(α_3^2, δ) Chart

In the course of the preceeding discussion a set of conditions for the various types of functions has been established in terms of α_3 and δ, parameters which may be readily calculated. The numerical values of these two parameters determine the Pearson curve appropriate to a particular distribution. The conditions for each type of curve are summarized in Table III.

An (α_3^2, δ) chart which gives visual presentation of these conditions and an automatic means for type identification is relatively easy to construct.

In addition to the lines, $\delta = -1$, $\delta = -1/2$, $\delta = 0$, $\delta = 2/5$, and $\alpha_3 = 0$, the chart contains only the curve:

(2-45) $$\alpha_3^2 = 4\delta(\delta+2)$$

TYPE	EQUATION	CONDITIONS	REMARKS
I	$y = C(t-r_1)^{m_1}(r_2-t)^{m_2}$	$\alpha_3 \neq 0$ $-1 < \delta < 0,\ \delta \neq -\frac{1}{2}$ $(2+3\delta)\alpha_3^2 \neq 4(1+2\delta)^2(2+\delta)$	Limited range, skew, usually bell-shaped but may be U-shaped or J-shaped. Range is (r_1, r_2).
II	$y = C(s^2-t^2)^m$	$\alpha_3 = 0$ $-1 < \delta < 0,\ \delta \neq -\frac{1}{2}$	Limited range, symmetrical, usually bell-shaped but U-shaped when $\alpha_4 < 1.8$. Range is $(-s, s)$.
III	$y = C(A+t)^{A^2-1} e^{-At}$	$\alpha_3 \neq 0$ $\delta = 0$	Unlimited range in one direction, skew; bell-shaped, but may be J-shaped. Range is $(-A, \infty)$.
VIII	$y = C(t-r_1)^{-2m}$	$\alpha_3 \neq 0$ $\delta < -\frac{1}{2}$ $(2+3\delta)\alpha_3^2 = 4(1+2\delta)^2(2+\delta)$	J-shaped. Range from infinite ordinate at r_1 to finite ordinate at $t = r_2$.
IX	$y = C(r_2-t)^{-2m}$	$\alpha_3 \neq 0$ $-\frac{1}{2} < \delta < 0$ $(2+3\delta)\alpha_3^2 = 4(1+2\delta)^2(2+\delta)$	Range from $t = r_2$ to infinite ordinate at r_1. J-shaped.
X	$y = e^{-t-1}$	$\alpha_3 \neq 0$ $\delta = 0$ $\alpha_3^2 = 4$	J-shaped with range $(-1, \infty)$.
XII	$y = C\left[\frac{r_2-t}{t-r_1}\right]^{m_2}$	$\delta \neq -\frac{1}{2}$	J-shaped with range (r_1, r_2).
IV.	$y = C\left[(t+r)^2+s^2\right]^{-m} e^{\frac{\nu\pi}{2}} e^{-\nu\tan^{-1}\left(\frac{t+r}{s}\right)}$	$\alpha_3 \neq 0$ $\delta > 0$ $\alpha_3^2 < 4\delta(\delta+2)$	Unlimited range, skew, bell-shaped.
V	$y = C(t+r)^{-2m} e^{-\frac{2r(m-1)}{t+r}}$	$\alpha_3 \neq 0$ $\delta > 0$ $\alpha_3^2 = 4\delta(\delta+2)$	Unlimited range in one direction, bell-shaped. Range is $(-r, \infty)$.
VII	$y = C(t^2+s^2)^{-m}$	$\alpha_3 = 0$ $\delta > 0$	Unlimited range, symmetrical, bell-shaped.
Normal	$y = Ce^{-\frac{t^2}{2}}$	$\alpha_3 = \delta = 0$	Unlimited range, symmetrical, bell-shaped.
VI	$y = Ct^{m_2}(t-\alpha)^{m_1}$	$\alpha_3 \neq 0$ $\delta > 0$ $\alpha_3^2 > 4\delta(\delta+2)$	Unlimited range in one direction, skew, bell-shaped but may be J-shaped.
XI	$y = C(t-r_2)^{-2m}$	$\alpha_3 \neq 0$ $0 < \delta < 2/5$ $(2+3\delta)\alpha_3^2 = 4(1+2\delta)^2(2+\delta)$	J-shaped with finite ordinate at $t = r_1$, Range is (r_1, ∞).

TABLE III Summary of Pearson System of Frequency Curves

on which the points corresponding to the Type V function lie, and the curve:

$$(2\text{-}46) \qquad (2 + 3\delta) = 4(1 + 2\delta)^2\ (2 + \delta)$$

on which the points corresponding to the distribution functions of Type VII, IX, X, and XI are found.[22]

Construction of $(\alpha_3^{\,2}, \delta)$ Chart

Point $\alpha_3^{\,2} = 0$, $\delta = 0$ satisfies the conditions for the normal curve and is the starting point for constructing the graph. The lines $\delta = -1$, $\delta = -1/2$, $\delta = 0$, $\delta = 2/5$, and $\alpha_3 = 0$, are easily constructed.

For the equation:

$$(2\text{-}45) \qquad \alpha_3^{\,2} = 4\delta(\delta+2)$$

δ	$\alpha_3^{\,2}$
0.	0.
.1	.84
.2	1.76
.3	2.73
.4	3.84

For the equation:

$$(2\text{-}46) \qquad (2+3\delta)\ \alpha_3^{\,2} = 4(1+2\delta)^2\ (2+\delta)$$

δ	$\alpha_3^{\,2}$
0.	4.
.2	6.634
.4	9.72
-.2	1.85
-.4	.32
-.5	0.
-.6	1.12

Note when $\delta = -2/3$, the expression $(2+3\delta) = 0$, therefore the line for this equation approaches $-2/3$ asymtopically.

Figure 2 presents visually the family of Pearson's curves in terms of α_3^2 and δ. It is a simple matter to determine the type of curve that fits any set of data merely by determining the α_3^2 and δ values for the data and then entering the chart with these values.

The subscript, B, on the chart refers to bell-shaped curves, the subscript, J, refers to J-shaped curves, and the subscript, U, to U-shaped curves.

The points for $\delta < -1$, correspond to no frequency functions, they fall in the "Impossible Area."

Pearson designated as heterotypic those members of his system for which the eighth movement failed to exist. (In such a case the standard deviation of the fourth moment in samples would be infinite.) This area is established by $\delta \geqq .4$.

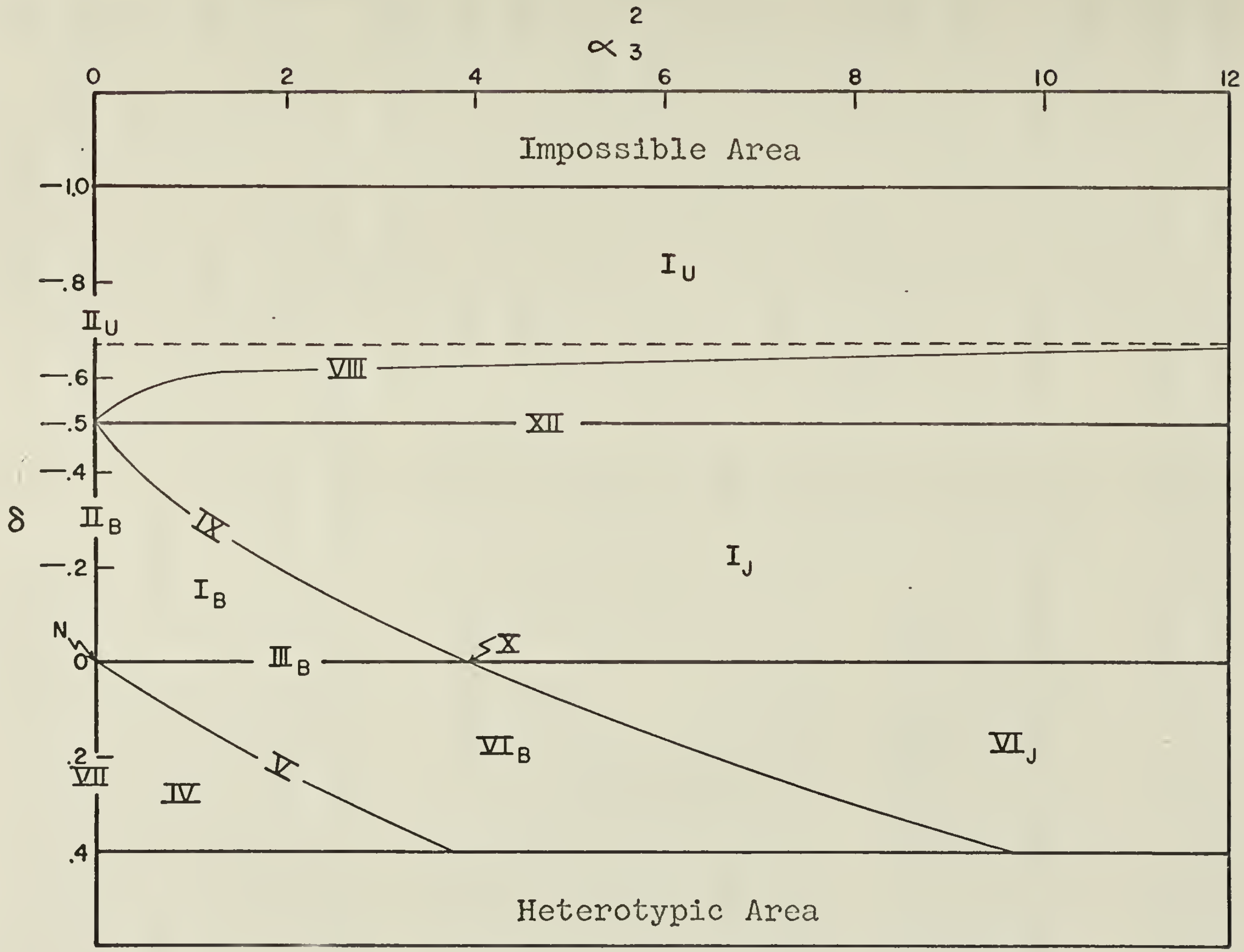


FIGURE 2 CHART OF PEARSON'S CURVES IN TERMS OF (α_3^2, δ)

CHAPTER III
APPLICATION OF PEARSON CURVES

Introduction

The brief theoretical discussion of the Pearson frequency curves having been presented it is now useful to examine how these curves may be applied to the field data under consideration.

From Table I there are eighteen sets of data available for the five fields. The Pearson system of frequency curves will be applied to find a good theoretical fit for each given observed distribution.

What is the value of frequency curves? A normal curve fitted to a given set of data is to determine whether or not the data are normally distributed. If the distribution is normal then use may be made of the extensive body of sampling theory applicable to normal populations. Comparatively little is known concerning sampling from non-normally distributed populations.

When data are distinguished as non-normal, there may be further advantages in fitting a non-normal curve to the data. Such a curve may serve to smooth the histogram and may thus permit a more accurate determination of the relative frequencies of the population from which the sample was taken. The identification of the distribution of the given data with a particular frequency curve may also serve to distinguish them from other data, the distribution of which is identified with a different frequency curve. These are

the two principle reasons for fitting non-normal frequency curves. Also, if a particular type of non-normal distribution occurs with sufficient frequency, this fact will serve as an incentive for the creation of the appropriate sampling theory.

Fitting the Data

The arithmetical labor involved in fitting a set of observed data to a frequency curve is lengthy and tedious. This may be a prime reason why there has not been greater utilization of theoretical curves in practice to obtain a description of the distribution of a given set of data. With the advent of high speed digital computers this objection to the heavy arithmetical work involved is lessened.

To illustrate the procedure involved in fitting a set of data to a Pearson frequency curve, the step by step calculations for one set of data, that of permeability for Field 3, is shown. The calculations for fitting the remaining sets of data were performed on a digital computer. The computer programs, written in Fortran language, with accompanying flow charts are shown in Appendix B. For each data set, the histogram and plotted curve follows the numerical calculations for that set. The normal curve has also been fitted to each set of data to illustrate the comparison with the non-normal curve that describes the data. In many cases the goodness of fit test for the normal curve indicates that the normal curve gives a poor fit for

the data, which is as one would expect for some of the more markedly skewed distributions. In a few cases though, the normal curve does give a satisfactory fit to the data, even though the normal fit is not as good as the frequency curve describing the data.

Procedure

The procedure used in fitting the various types of curves is:

1. Arrange the data in an array. Tabulate the data using convenient class intervals.
2. Calculate the first four moments about a convenient vertical.
3. Transfer the moments to the centroid vertical or vertical through the mean.
4. Apply Sheppard's corrections[1] to the moments if there is high contact at both ends of the curve.
5. Calculate α_3^2, α_4, and δ.
6. Locate the mean $\overline{X}$ and the mode $\overline{\mu}$.
7. Determine by use of the (α_3^2, δ) chart what type of curve should be used.
8. Calculate the constants for the equation of the curve.
9. Calculate the theoretical frequencies at the mid-point of each class interval.
10. Plot the histogram.
11. Plot the theoretical curve constructing the mid-ordinates at the middle of each class interval.

12. Calculate the area graduation for each interval. Test for goodness of fit of theory to observation.
13. Fit the normal curve to the data for comparison and test for goodness of fit of the normal curve.

Sample Calculations

The permeability data for Field 3 consists of 195 observations obtained from seven core analyses from seven wells. These data were divided into 14 class intervals of 10 millidarcies each. Table IV shows a convenient method of calculating the first four moments for the core data. Following Table IV are the calculations needed to fit the frequency curve to the data.

The fitting of a set of data to a Pearson curve is based on the method of moments where the r^{th} moment around an arbitrary origin is:[2]

(3-1) $$\mu_r' = \int_{-\infty}^{\infty} X^r \, dF(X)$$

The zero-th moment about the origin, μ_o', always exists and is equal to one.

The r^{th} moment around the mean is:

(3-2) $$\mu_r = \int_{-\infty}^{\infty} (X - \mu_1') \, dF(X)$$

If the moments around an arbitrary origin are known, the following formula is used to find the moments about the mean:[3]

(3-3) $$\mu_r = \mu_r' - r\,\mu_{r-1}'\,\mu_1' + \frac{r(r-1)}{2\,!}\,\mu_{r-2}'\,\mu_1'^2 + \ldots + (-1)^r\,\mu_1'^r$$

Thus:

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\,\mu_2' + 2\mu_1'^3$$

$$\mu_4 = \mu_4' - 4\mu_1'\,\mu_3' + 6\mu_1'^2\,\mu_2' - 3\mu_1'^4$$

For ease of hand calculation and simplicity of notation, the scale for the independent variable of the frequency distribution, (i.e., porosity, permeability, or saturation) was transformed to a notation wherein the interval containing the arbitrarily chosen mid-point is numbered zero and intervals on either side are numbered serially. The negative values were assigned to that side of the distribution which contained the mode. This convention is in keeping with that adopted for the sign of "a" in the derivation of the Pearson system of frequency curves. An additional transformation of the independent variable is made to correspond to standard statistical notation. This transformation is the introduction of the standard unit t. Thus when the histogram intervals are numbered serially,

(3-4) $$t = \frac{x_i - \mu_1'}{\sigma}$$

where x_i takes on the values 1, 2, 3, 4, etc. at interval mid-points.

The equation relating t and the original physical property, q, is:

$$q = t\ \sigma\ \Delta q + \overline{q} \tag{3-5}$$

where $\overline{q}$ is the mean value of q and Δq is the interval of q used to construct the histogram on the q scale.

The following calculations starting with Table IV are for the permeability observations for Field 3. For each of the other data sets only the resulting constants and equations are shown followed by the graph of the frequency curve and normal curve.

Taking moments about the arbitrary mid-ordinate 45.0 in terms of the transformed variable x with its corresponding mid-point, x = 0, the following moments are calculated:

$$\mu_1' = \frac{\Sigma xf}{\Sigma f} = \frac{-90}{195} = -.46153846 \tag{3-6}$$

$$\mu_2' = \frac{\Sigma x^2 f}{\Sigma f} = \frac{1260}{195} = 6.4615385 \tag{3-7}$$

$$\mu_3' = \frac{\Sigma x^3 f}{\Sigma f} = \frac{960}{195} = 131.2000 \tag{3-8}$$

TABLE IV

CALCULATION OF THE FIRST FOUR MOMENTS OF THE DISTRIBUTION OF PERMEABILITY--FIELD 3

Permeability Millidarcies	Mid-Value of Interval	Frequency f	x	xf	x^2	x^2f	x^3	x^3f	x^4	x^4f
0 - 10	5	15	-4	-60	16	240	-64	-960	256	3,840
10.01- 20	15	29	-3	-87	9	261	-27	-783	81	2,349
20.01- 30	25	34	-2	-68	4	136	- 8	-272	16	544
30.01- 40	35	32	-1	-32	1	32	- 1	- 32	1	32
40.01- 50	45	24	0	0	0	0	0	0	0	0
50.01- 60	55	20	1	20	1	20	1	20	1	20
60.01- 70	65	16	2	32	4	64	8	128	16	256
70.01- 80	75	11	3	33	9	99	29	297	81	891
80.01- 90	85	8	4	32	16	128	64	512	256	2,048
90.01-100	95	2	5	10	25	50	125	250	625	1,250
100.01-110	105	1	6	6	36	36	216	216	1296	1,296
110.01-120	115	1	7	7	49	49	343	343	2401	2,401
120.01-130	125	1	8	8	64	64	512	512	4096	4,096
130.01-140	135	1	9	9	81	81	729	729	6561	6,561
Totals of Terms	-	195	-	-90	-	1260	-	960		25,584

Assuming 45.0 as the mid-point, the following are calculated:

$\Sigma xf = -90$

$\Sigma x^2 f = 1,260$

$\Sigma x^3 f = 960$

$\Sigma x^4 f = 25,584$

$$\overline{X} = \frac{7966}{195} = 40.85$$

And, for moments about the mean,

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2 = 6.4615385 - (-.46153846)^2$

$\mu_2 = 6.2485208$

$\mu_3 = \mu_3' - 3\mu_1' \mu_2' + 2\mu_1'^3$

$\mu_3 = 4.9230769 - 3(-.46153846)(6.461538) + 2(-.46153846)^3$

$\mu_3 = 13.673191$

$\mu_4 = \mu_4' - 4\mu_1' \mu_3' + 6\mu_1'^2 \mu_2' - 3\mu_1'^4$

$\mu_4 = 148.41116$

And the standard deviation for grouped data,

(3-9) $$\sigma = \sqrt{\mu_2} = \sqrt{6.1651875} = 2.4997035$$

Hence Sheppard's corrections are:

(3-10) $$X_2 = \mu_2 - \frac{1}{12} = 6.2485208 - .083333$$

$$X_2 = 6.1651875$$

(3-11) $$X_3 = \mu_3 = 13.673191$$

(3-12) $$X_4 = \mu_4 - \frac{1}{2}\mu_2 + \frac{7}{240}$$

$$X_4 = 148.41116 - \frac{1}{2}(6.2485208) + .029167$$

$$X_4 = 145.31607$$

Thus,

$$\alpha_3^{\ 2} = \frac{X_3^{\ 2}}{X_2^{\ 3}} = \frac{(13.673191)^2}{(6.1651875)^3} = .79781261$$

$$\alpha_4 = \frac{X_4}{X_2{}^2} = \frac{145.31607}{(6.1651875)^2} = \frac{145.31607}{38.0095364} = 3.8231476$$

$$\delta = \frac{2\,\alpha_4 - 3\,\alpha_3{}^2 - 6}{\alpha_4 + 3}$$

$$\delta = \frac{2(3.8231476) - 3(.79781261) - 6}{(3.8231476) + 3}$$

$$\delta = -.10950116$$

$$D = \alpha_3{}^2 - 4\delta(\delta + 2)$$

$$D = .79781261 - 4(-.10950116)(1.8904988)$$

$$D = 1.6258600$$

Then entering the $(\alpha_3{}^2, \delta)$ chart with:

$$\alpha_3{}^2 = .79781261 \text{ and } \delta = -.10950116$$

it is observed that the frequency distribution of the permeability for Field 3 may be fitted by a Pearson Type I_B curve.

The equation of the curve is:

$$Y = C(t - r_1)^{m_1} \;\; (r_2 - t)^{m_2}$$

where:

$$C = \frac{N\,\Gamma(m_1 + m_2 + 2)}{(r_2 - r_1)^{m_1 + m_2 + 1}\;\Gamma(m_1 + m_2)\;\Gamma(m_2 + 1)}$$

where N is the sum of the frequency for all class intervals and:

$$r_1 = \frac{-\alpha_3 + \sqrt{D}}{2\,\delta}$$

$$r_1 = \frac{-.89320367 + \sqrt{1.625860}}{2(-.10950116)}$$

$$r_1 = -1.7437633$$

$$r_2 = \frac{-\alpha_3 - \sqrt{D}}{2\ \delta}$$

$$r_2 = 9.9007883$$

$$m_1 = -\left(\frac{1+\delta}{\delta}\right)\left(1 - \frac{\alpha_3}{\sqrt{D}}\right) - 1$$

$$m_1 = -\left(\frac{1 - .10950116}{-.10950116}\right)\left(1 - \frac{.89320367}{1.2750919}\right) - 1$$

$$m_1 = 1.4356194$$

$$m_2 = -\left(\frac{1+\delta}{\delta}\right)\left(1 + \frac{\alpha_3}{\sqrt{D}}\right) - 1$$

$$m_2 = 12.829027$$

And:

$$C = \frac{195\ \Gamma(1.4356194) + 12.829027 + 2)}{(9.9007883 + 1.7437633)^{15.264646}\ \Gamma 14.264646\ \Gamma 13.829027}$$

$$C = \frac{195\ \Gamma 16.2646464}{(11.6445516)^{15.26}\ \Gamma 14.264646\ \Gamma 13.829027}$$

$$C = 2.211844 \times 10^{-12}$$

The Gamma function, defined as:[4]

(3-12) $$\Gamma(x) = \int_0^\infty e^{-t}\ t^{(x-1)}\ dt,\ x > 0$$

was calculated in two ways, depending upon the value of x.

For values of x less than 9.0, the Gamma function was obtained from the recursion formula:[5]

$$\Gamma(x+1) = x\,\Gamma(x) \tag{3-13}$$

which is useful for expressing $\Gamma(x)$ as a function of some value of $\Gamma(a)$ where (a) is a value between 1. and 2. The value of $\Gamma(a)$ was determined by a table look-up and interpolation using tabulated Gamma functions.[6]

For values of x above 9.0, $\Gamma(x)$ was computed from the Stirling formula[7] for $\log \Gamma(x)$.

$$\begin{aligned} \log \Gamma(x) &= -.43429448(x) + (x - .5)\log_{10} x \\ &+ \log_{10}\left(1 + \frac{1}{12x}\right) + .39908993 \end{aligned} \tag{3-14}$$

The equation of the curve is:

$$Y = (2.211844\text{x}10^{-12})(t+1.7437633)^{1.4356194}\,(9.90079-t)^{12.829}$$

The range of the curve can be computed from the above equation.

The range of the curve is defined as the upper and lower values of the independent variable between which all positive frequencies exist. These theoretical limits are determined by seeking the values of t for which Y = 0. By inspection of the above equation, Y is zero when t is -1.7437633 and +9.90070. These values of t correspond to permeabilities of -2.65 md and 288.35 respectively as determined from the equation.

$$q = \bar{q} + t\,\sigma\,\Delta q$$

where t is one of the limiting values given above, σ the standard deviation in units of x, Δ q is the class interval in units of q, and $\overline{q}$ is the mean value of q. Because negative values of permeability are not physically meaningful, the limits can be looked upon as zero and 288.35 millidarcys. The existence of the negative lower limit should present no more interpretational difficulties than those presented by the known limits of any normal frequency curve which are by definition, $-\infty$ and $+\infty$.

For calculation of graduation (mid-ordinates), the following arrangement is convenient:

(1) Mid-point	(2) Distance from origin	(3) $t = (\frac{2}{\sigma})$	(4) (3) - r_1	(5) log(4)	(6) r_2-(3)
65	2.46153846	.98473219	2.7284955	.43592	8.9160561

(7) Log(6)	(8) m_1x(5)	(9) m_2x(7)	(10) (8)+(9)+LogC=LogY_x	(11) Y_x	(12) Area of Interval
.95017	.624807	12.18973	1.16038	14.467	14.512

The area of the interval is found by means of Simpson's Rule:[8]

(3-15) $$\int_0^1 f(x)dx = 1/6(Y_o + 4Y_{1/2} + Y_1)$$

By use of the above calculating techniques, the values of Y_x (Graduation-mid-ordinates) and Graduation (Areas) was found for each interval. Additionally the values of Y_{normal}

(Graduation-mid-ordinates) was found for each mid-point by substituting the appropriate value of t for each mid-point into the equation:

$$Y = C_N \, e^{-\frac{t^2}{2}}$$

where:

$$C_N = \frac{N}{\sqrt{2\pi}\ \sigma}$$

Thus:

$$Y_{65} = \frac{195}{\sqrt{2\pi}\ \sigma} \, e^{\frac{-(.9847)^2}{2}} = 19.164$$

The following summary for Field 3 permeability data is presented: Type I_B Curve

$$Y = (2.21184 \times 10^{-12})(t + 1.7437633)^{1.4356194}\ (9.90079-t)^{12.829}$$

Permeability Midpoint (md)	Frequency	Graduation (mid-ordinates)	Graduation (Areas)	Normal Curve (mid-ordinates)
5	15	14.73	14.48	11.43
15	29	29.14	28.72	18.58
25	34	33.85	33.57	25.75
35	32	31.81	31.68	30.41
45	24	26.47	26.44	30.60
55	20	20.23	20.25	26.23
65	16	14.47	14.51	19.16
75	11	9.77	9.82	11.93
85	8	6.26	6.31	6.33
95	2	3.82	3.86	2.86
105	1	2.22	2.24	1.10
115	1	1.22	1.24	.36
125	1	.64	.65	.10
135	1	.32	.32	.02
	195	194.95	194.09	184.86

Goodness of Fit

The Chi Square (χ^2) goodness of fit test for the above curve may be calculated by the following method:

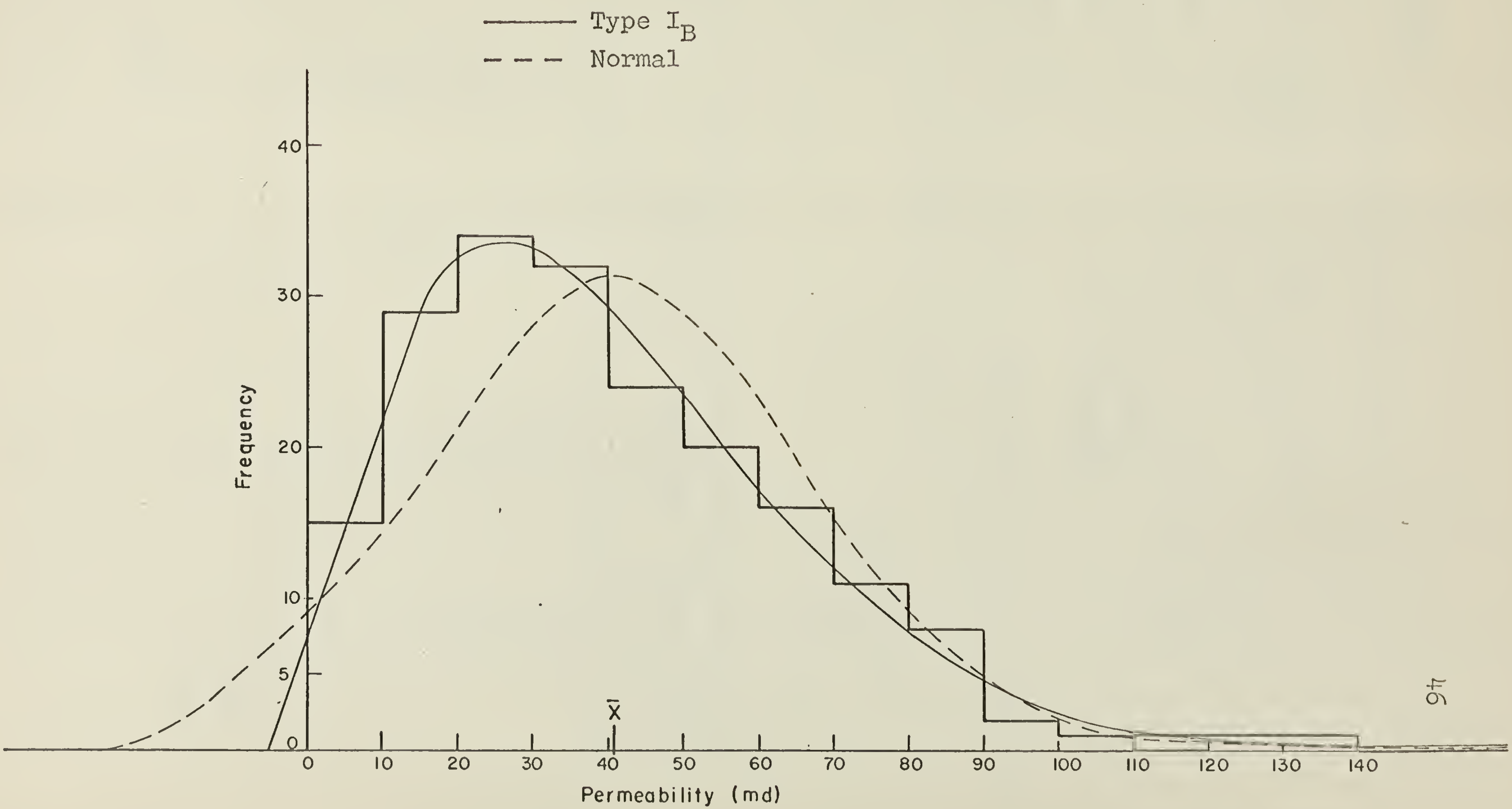
Chi-Square "Goodness of Fit"

f	Graduated F	f - F	$(f - F)^2$	$\frac{(f - F)^2}{F}$
15	14.5	.5	.25	.018
29	28.7	.3	.09	.003
34	33.6	.4	.16	.006
32	31.7	.3	.09	.003
24	26.4	-2.4	5.76	.225
20	20.3	- .3	.09	.003
16	14.5	1.5	2.25	.152
11	9.8	1.2	1.44	.142
8	6.3	1.7	2.89	.444
2	3.8	-1.9	3.61	.893
1	2.2	-1.2	1.44	.689
1	1.2	- .2	.04	.046
1	.7	.3	.09	.188
1	.3	.7	.21	.700
	194.1			3.522

The number of degrees of freedom for this curve is n - 1 = 13, where n is the number of class intervals. Entering a χ^2 table[9] with 13 degrees of freedom and $\chi^2 = 3.522$ an approximate value of p = .99 is found. This indicates that there are 99 chances out of 100 that differences as large as those found could have arisen due to chance or sampling variation. Therefore, we can conclude that the Type I_B gives an exceptionally good fit to the permeability data. For testing the goodness of fit the significance level of p is usually taken as either .05 or .01.[10]

FIGURE 3 TYPE I_B PERMEABILITY FIELD 3

RANGE (-2.65, 288.4)



Numerical Results

Permeability (K)--Field 1

Data Range (.2, 439 md.)   Mean = 63.26 md.

Class interval = 40 md. σ = 60.8 md.

$\alpha_3^{\ 2} = 3.0327404$ $\delta = -.2034506$ Type I_J curve

$r_1 = -.93047735$ $r_2 = 9.4901735$

$m_1 = -.3008110$ $m_2 = 6.1312036$

$C = 2.8551007 \times 10^{-4}$

$$Y = (2.8551007\text{x}10^{-4})(t+.93047735)^{-.30081100}\ (9.4901735-t)^{6.1312036}$$

Curve Range (6.7, 490 md.)

I_J Curve

$\chi^2 = 17.48$

Fit is good at .05 level

Normal Curve

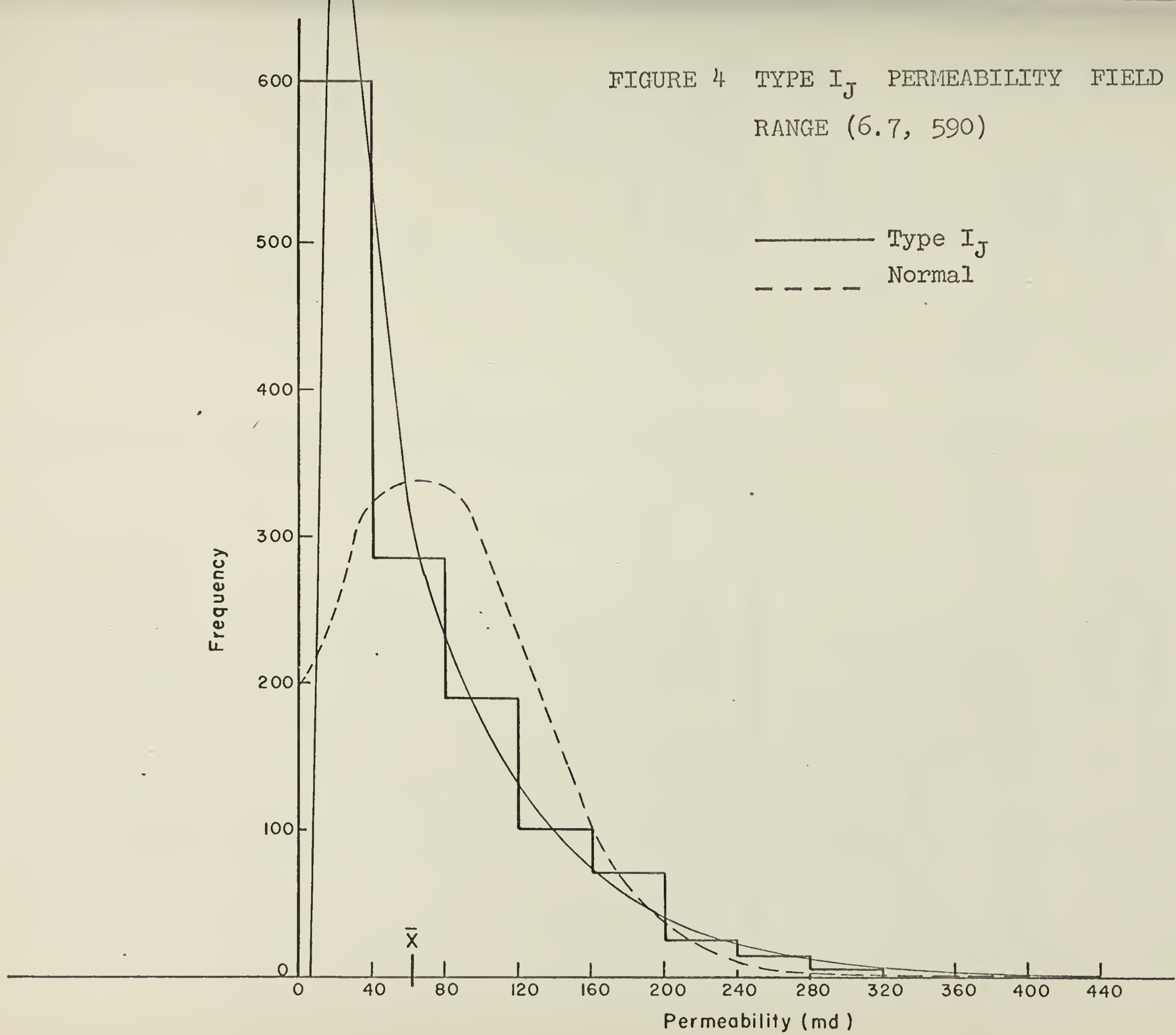
$\chi^2 = 968.2$

Fit is not good

Permeability Midpoint (md.)	Frequency	Graduation (mid-ordinate)	Graduation (Areas)	Normal Curve (mid-ordinate)
20	610	786.1	600.8	253.3
60	284	320.9	329.7	338.7
100	189	174.1	176.9	294.6
140	101	97.7	99.1	166.7
180	71	54.4	55.1	61.3
220	25	29.4	29.9	14.7
260	13	15.3	15.5	22.8
300	5	7.5	7.6	.2
340	1	3.4	3.5	.02
380	3	1.4	1.4	.0007
420	1	.5	.5	.00002
	1303			

FIGURE 4 TYPE I_J PERMEABILITY FIELD 1

RANGE (6.7, 590)



Porosity (ϕ)--Field 1

Data Range (5.4, 25.5%) Mean = 19.43%

Class interval = 2.5% $\sigma = 3.65$

$\alpha_3{}^2 = .72488482$ $\delta = -.29374565$ Type I_B curve

$r_1 = -1.3630479$ $r_2 = 4.2614796$

$m_1 = .16531900$ $m_2 = 2.6432928$

$C = 3.7594063$

$$Y = (3.7594063)(t+1.3630479)^{1.6531900}(4.2614796-t)^{2.6432928}$$

Curve Range (3.83, 24.0%)

I_B Curve

$\chi^2 = 19.841246$

Fit is good at .02 level

Normal Curve

$\chi^2 = 160.8$

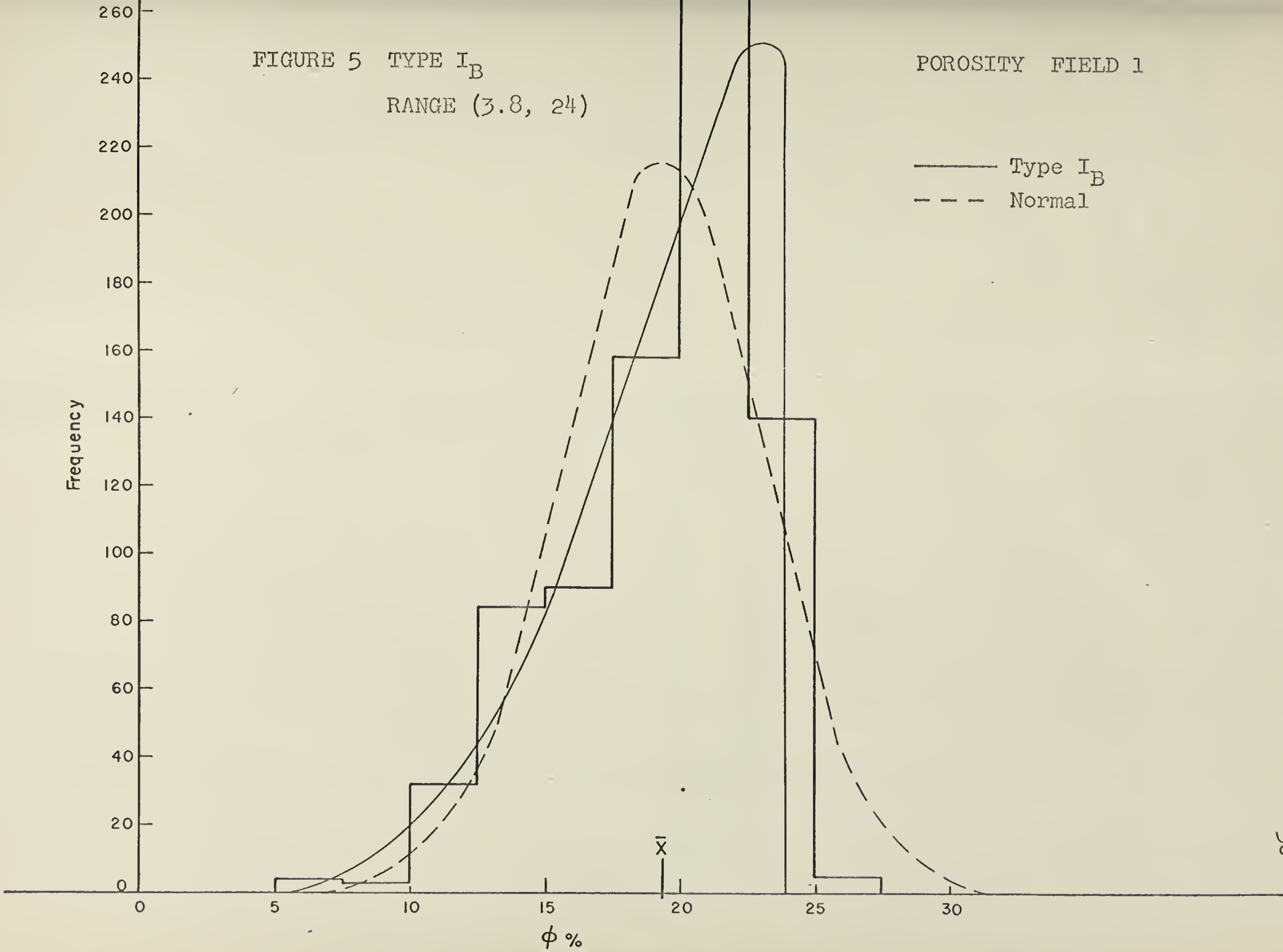
Fit is not good

Porosity Midpoint (%)	Frequency	Graduation (mid-ordinate)	Graduation (Areas)	Normal Curve (mid-ordinate)
6.25	4	1.8	2.1	.4
8.75	3	10.9	11.4	3.3
11.25	32	31.0	31.5	18.8
13.75	84	64.0	64.5	67.4
16.25	90	110.3	110.8	151.6
18.75	158	168.0	168.2	213.6
21.25	278	228.8	227.7	188.8
23.75	140	246.9	206.6	104.6
26.25	5	0.	0.	36.4
	794			

FIGURE 5 TYPE I_B

RANGE (3.8, 24)

POROSITY FIELD 1



Oil Saturation (S_o)--Field 1

Data Range (7.0, 73.0%) Mean = 38.89%

Class interval = 5% σ = 11.05

α_3^2 = .020572560 δ = .027215431 Normal Type curve

C = 142.75

$$Y = 142.75\, e^{-\frac{t^2}{2}}$$

Curve Range ($-\infty$, ∞)

Normal

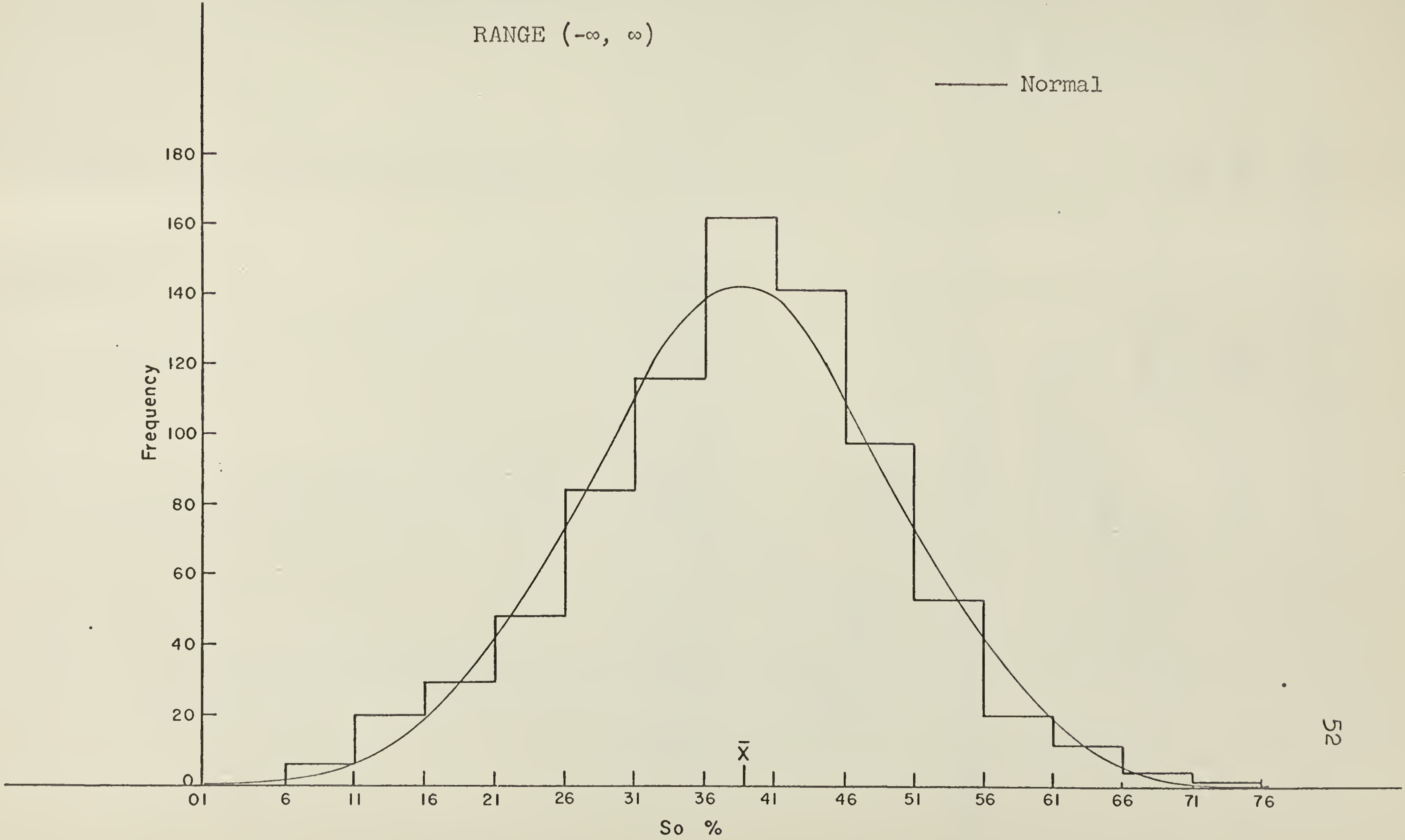
χ^2 = 16.568630

Fit is good at .25 level

S_o Midpoint (%)	Frequency	Normal Mid-ordinates
8.5	6	3.7
13.5	20	11.3
18.5	29	28.3
23.5	48	57.7
28.5	84	95.8
33.5	116	129.7
38.5	162	143.2
43.5	141	128.8
48.5	98	94.5
53.5	53	56.5
58.5	20	27.6
63.5	12	11.0
68.5	4	3.5
73.5	1	.9
	794	

FIGURE 6 NORMAL   OIL SATURATION FIELD 1

RANGE $(-\infty, \infty)$



Water Saturation (S_w)--Field 1

Data Range (15, 89%) Mean = 40.96%

Class interval = 5% σ = 12.15

$\alpha_3^2 = .72375019$ $\delta = -.10356269$ Type I_B curve

$r_1 = -1.8241159$ $r_2 = 10.038806$

$m_1 = 1.6619956$ $m_2 = 13.649978$

$C = 9.5915501 \times 10^{-13}$

$$Y = (9.5915501 \times 10^{-13})(t+1.8241159)^{1.6619956}\ (10.938806-t)^{13.649978}$$

Curve Range (18.9, 171)

I_B Curve Normal Curve

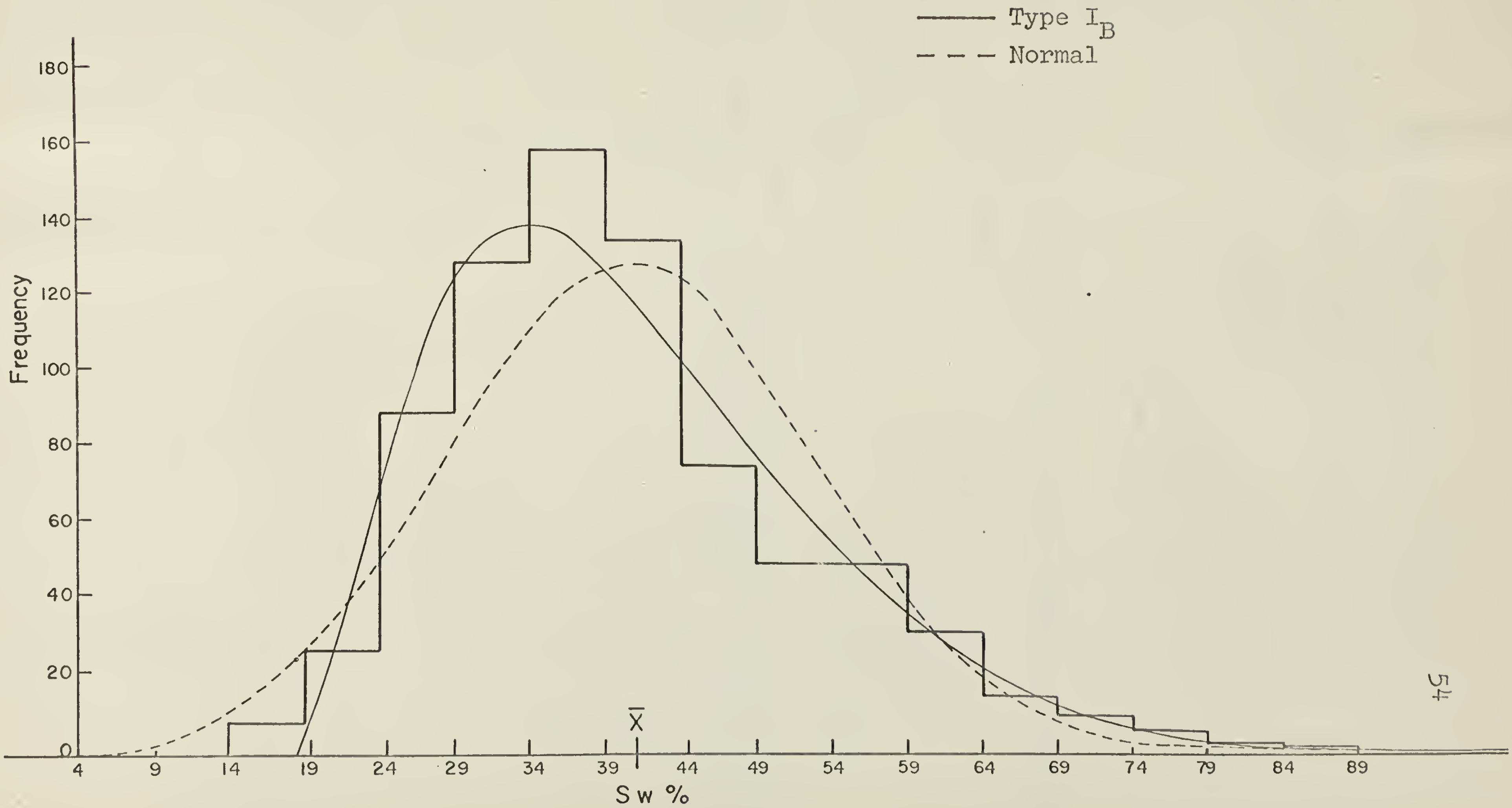
$\chi^2 = 10.875972$ $\chi^2 = 95.56815$

Fit is good at .80 level Fit is not good

S_w Midpoint (%)	Frequency	Graduation Mid-ordinates	Graduation (Areas)	Normal Curve Mid-ordinates
16.5	8	0.	.6	18.6
21.5	27	35.4	36.3	38.6
26.5	89	103.2	101.6	67.3
31.5	130	136.5	135.1	99.2
36.5	160	136.8	136.0	123.4
41.5	136	117.8	117.5	129.8
46.5	76	91.6	91.7	115.2
51.5	51	65.9	66.1	86.3
56.5	49	44.3	44.6	54.6
61.5	32	28.1	28.3	29.2
66.5	15	16.9	17.0	13.2
71.5	10	9.6	9.7	5.0
76.5	6	5.1	5.2	1.6
81.5	3	2.6	2.7	.4
86.5	2	1.2	1.3	.1
	794			

FIGURE 7 TYPE I_B WATER SATURATION FIELD 1

RANGE (18.9, 171)



Permeability (K)--Field 2

Data Range (.16, 381.0 md.) Mean = 49.8 md.

Class interval = 30 md. σ = 53.4 md.

$\alpha_3^{\ 2} = 5.4376534$ $\delta = -.22554678$ Type I_J curve

$r_1 = -.71192903$ $r_2 = 11.050699$

$m_1 = -.58435670$ $m_2 = 5.4516965$

$C = 1.8509951 \times 10^{-4}$

$$Y = (1.8509951\text{x}10^{-4})(t+.71192903)^{-.58435670}\ (11.050699-t)^{5.4516965}$$

Curve Range (11.9, 636.8 md.)

I_J Curve

$\chi^2 = 24.96777$

Fit is good at .02 level

Normal Curve

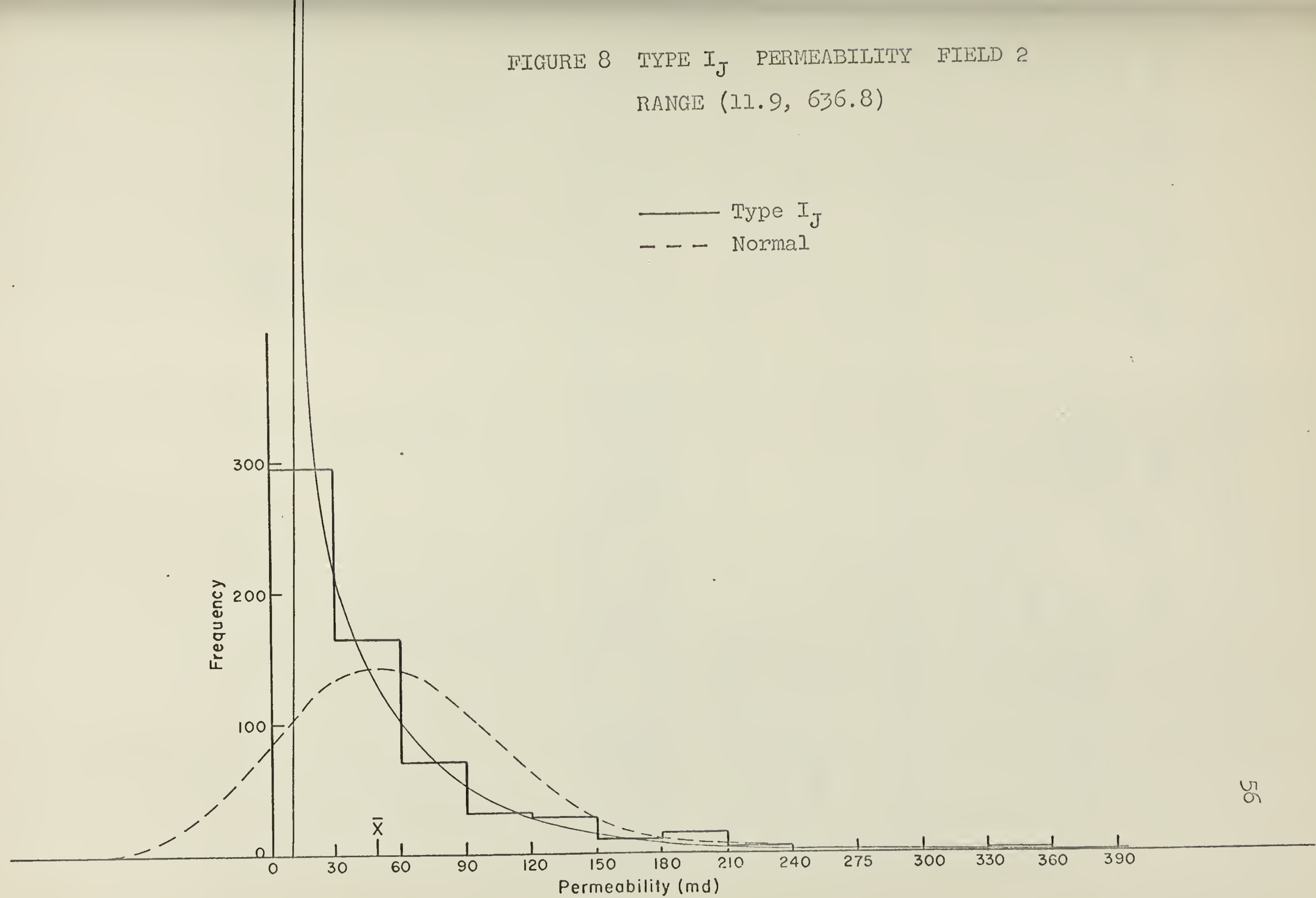
$\chi^2 = 486.497$

Fit is not good

Permeability Mid-point (md.)	Frequency	Graduation (mid-ordinates)	Graduation (Areas)	Normal Curve (mid-ordinates)
15	295	1080.9	757.2	111.6
45	164	131.0	139.1	140.1
75	71	66.8	68.1	128.4
105	32	39.4	39.8	85.7
135	28	24.4	24.6	41.8
165	12	15.3	15.5	14.8
195	17	9.7	9.8	3.8
225	5	6.1	6.1	.7
255	2	3.8	3.8	.1
285	1	2.2	2.3	.01
315	0	1.3	1.3	.0007
345	2	.7	.7	.00004
375	1	.4	.4	.000001
	630			

FIGURE 8 TYPE I_J PERMEABILITY FIELD 2

RANGE (11.9, 636.8)



Porosity (ϕ)--Field 2

Data Range (8.1, 24.8) Mean = 19.72%

Class interval = 2% σ = 2.34%

$\alpha_3^{\ 2} = 1.3635387$ $\delta = .14055615$ Type VI_B curve

$m_1 = 4.5488532$ $m_2 = -22.778043$

$r_1 = -1.9698380$ $r_2 = -7.7311868$

$C = 20.968745$

$$Y = (20.968745)(t+7.7311868)^{-22.778043}\ (t+1.9698380)^{4.5488532}$$

Curve Range ($-\infty$, 24.33)

VI_B Curve

$\chi^2 = 15.436512$

Fit is good at .05 level

Normal Curve

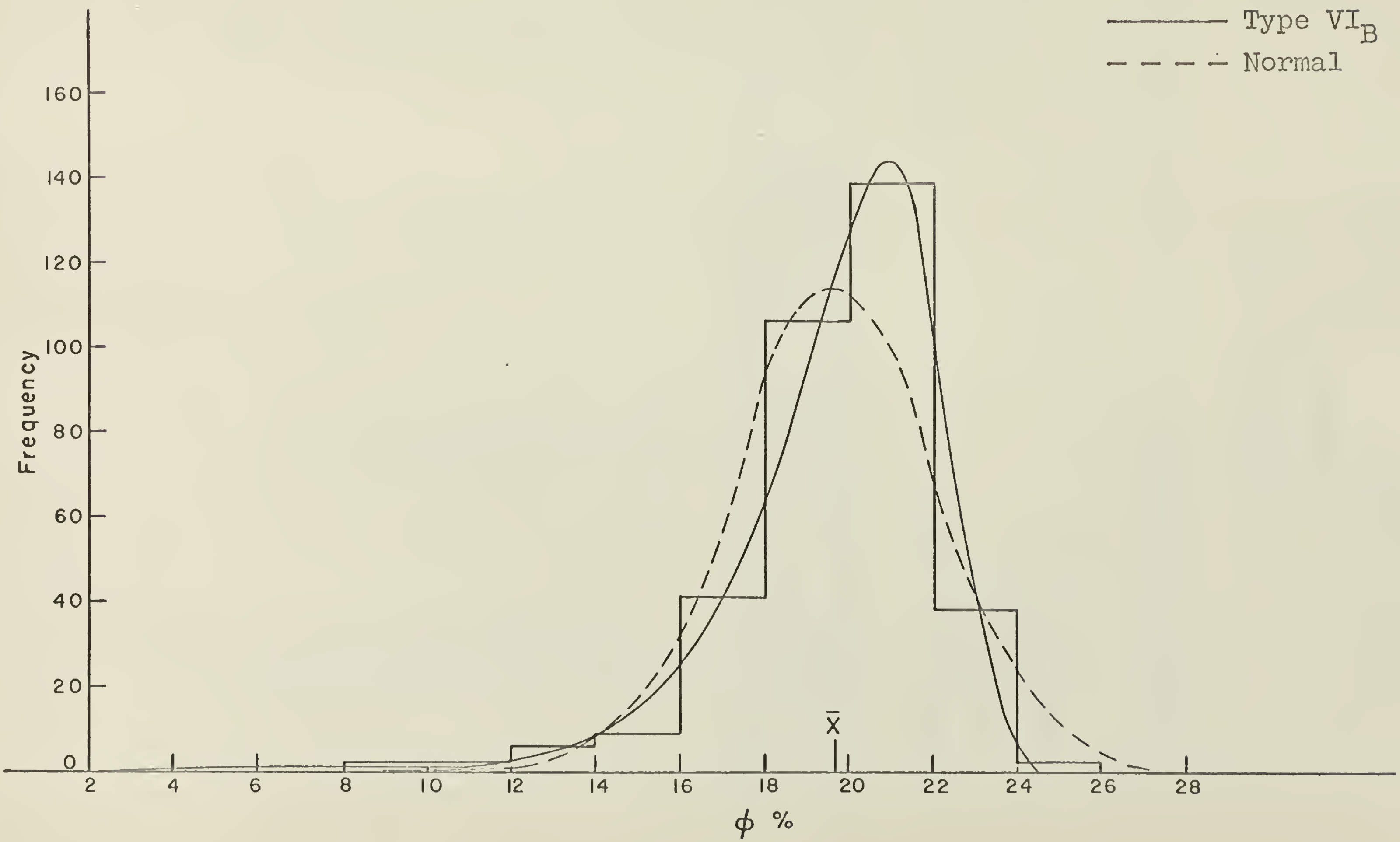
$\chi^2 = 71.647549$

Fit is not good

Porosity Midpoint (%)	Frequency	Graduation (mid-ordinates)	Graduation Areas	Normal Curve (mid-ordinates)
9	2	.7	.7	.003
11	2	1.9	2.0	.1
13	6	5.4	5.6	1.9
15	9	15.0	15.7	15.4
17	41	40.4	41.7	59.6
19	106	94.4	94.8	111.6
21	139	144.2	136.7	101.3
23	38	41.4	47.0	44.5
25	2	0.	.1	9.5
	345			

FIGURE 9 TYPE VI_B POROSITY FIELD 2

RANGE $(-\infty, 24.3)$



Oil Saturation (S_o)--Field 2

Data Range (9, 79%) Mean = 49.38%

Class interval = 7% σ = 14.7

α_3^2 = .6911813 δ = -.30506016 Type I_B curve

r_1 = -1.3600193 r_2 = 4.0852955

m_1 = .13792540 m_2 = 2.4181582

C = 1.7498843

$$Y = (1.7498843)(t+1.3600193)^{1.379254} (4.0852955-t)^{2.4181582}$$

Curve Range (-10.7, 69.4)

I_B Curve

χ^2 = 21.737241

Fit is good at .01 level

Normal Curve

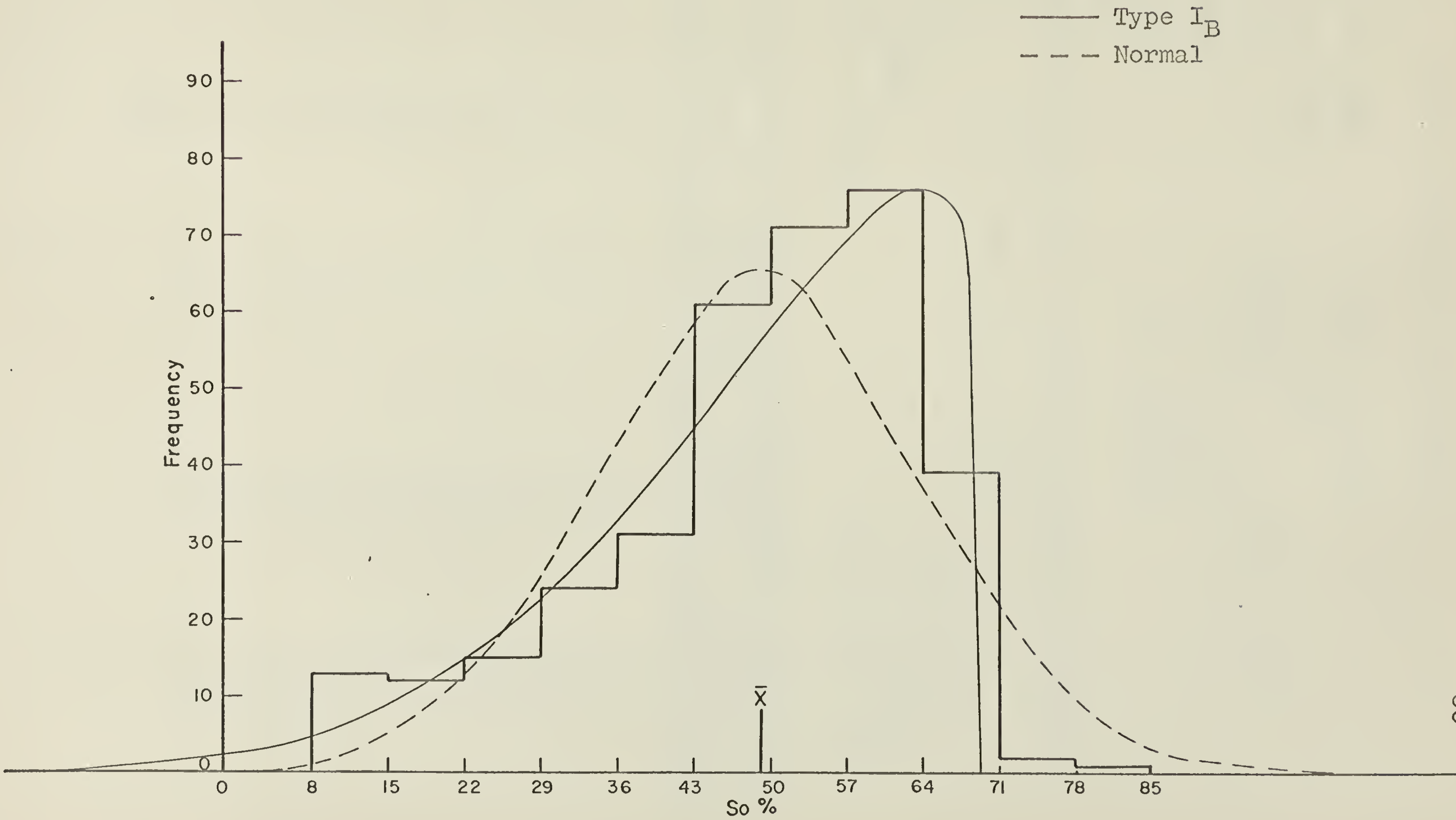
χ^2 = 90.32

Fit is not good

S_o Midpoint (%)	Frequency	Graduation Mid-ordinate	Graduation Areas	Normal (mid-ordinates)
11.5	13	6.1	6.1	2.6
18.5	12	11.4	11.5	7.8
25.5	15	18.6	18.7	18.6
32.5	24	27.6	27.7	35.3
39.5	31	38.4	38.5	53.5
46.5	61	50.6	50.6	64.7
53.5	71	63.4	63.3	62.3
60.5	76	74.7	74.3	47.9
67.5	39	72.8	61.4	29.4
74.5	2	0.	0.	14.3
81.5	1	0.	0.	5.6
	345			

FIGURE 10 TYPE I_B OIL SATURATION FIELD 2

RANGE (-10.7, 69.4)



Water Saturation (S_w)--Field 2

Data Range (11, 85%) Mean = 38.65%

Class interval = 6% σ = 14.82

$\alpha_3^2 = 1.0579891$ $\delta = -.32958992$ Type I_J curve

$r_1 = -1.1787589$ $r_2 = 4.2995634$

$m_1 = -.12466560$ $m_2 = 2.1928124$

$C = 1.8860123$

$$Y = (1.8860123)(t+1.1787589)^{-.12466560}\,(4.2995634-t)^{2.1928124}$$

Curve Range (21.2, 102.3)

I_J Curve

$\chi^2 = 7.3586264$

Fit is good at .75 level

Normal Curve

$\chi^2 = 98.996253$

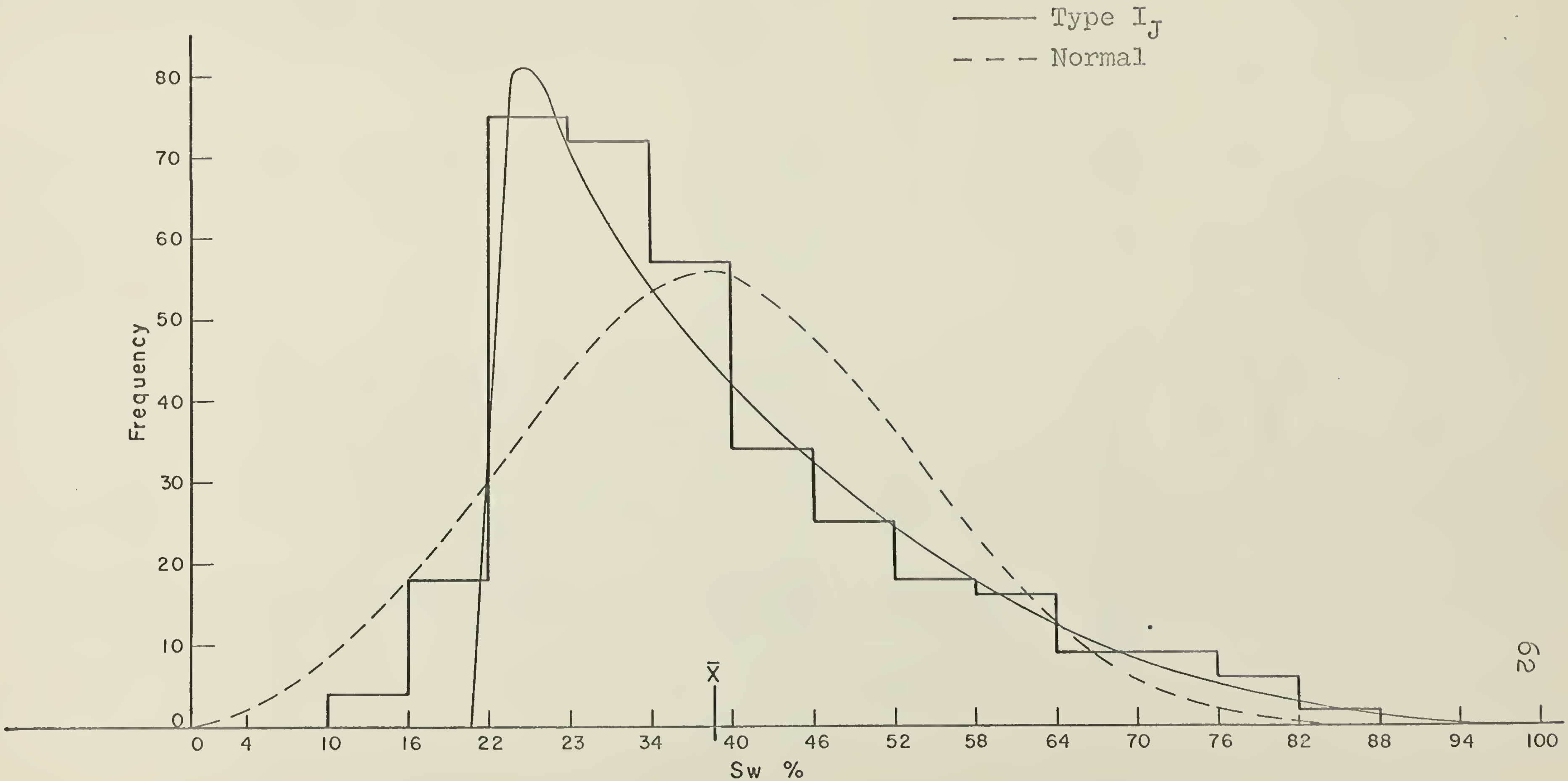
Fit is not good

S_w Mid-point (%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal (mid-ordinates)
13	4	0.	0.	13.2
19	18	0.	17.1	24.2
25	75	81.3	82.9	37.5
31	72	61.0	61.3	49.5
37	57	47.5	47.6	55.5
43	34	36.9	37.0	52.8
49	25	28.3	28.4	42.6
55	18	21.2	21.3	29.2
61	16	15.4	15.5	17.0
67	9	10.7	10.7	8.4
73	9	7.0	7.0	3.5
79	6	4.1	4.2	1.3
85	2	2.1	2.1	.4
	345			

FIGURE 11 TYPE I_J

RANGE (21.2, 102.3)

WATER SATURATION FIELD 2



Porosity (ϕ)--Field 3

Data Range (10.2, 23.1%) Mean = 18.06%

Class interval = 1.5% σ = 2.46

$\alpha_3^{\,2}$ = .58195415 δ = -.25091835 Type I_B curve

r_1 = -1.526424 r_2 = 4.5666935

m_1 = .4957617 m_2 = 3.4749591

C = .11555541

$$Y = (.11555541)(t+1.526424)^{4.957617}\,(4.5666935-t)^{3.4749591}$$

Curve Range (6.7, 21.8)

I_B Curve

χ^2 = 17.752871

Fit is good at .05 level

Normal Curve

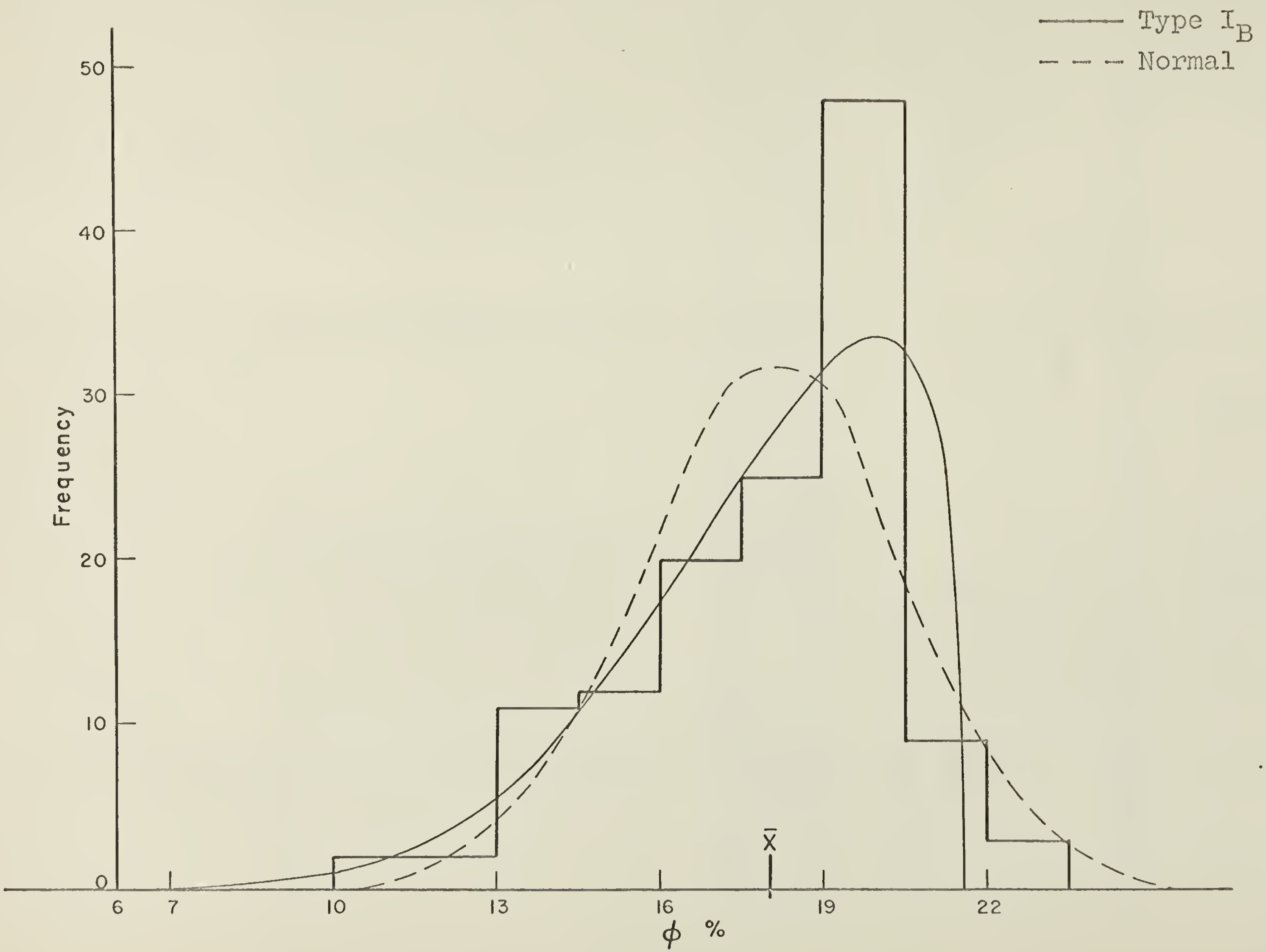
χ^2 = 36.031458

Fit is not good

Porosity Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal (mid-ordinates)
10.75	2	1.3	1.3	.4
12.25	2	3.6	3.7	2.1
13.75	11	7.8	7.8	7.1
15.25	12	13.8	13.8	17.0
16.75	20	21.3	21.3	28.0
18.25	25	29.0	28.9	31.9
19.75	48	33.8	33.3	25.1
21.25	9	25.5	22.5	13.6
22.75	3	0.	0.	5.1
	132			

FIGURE 12 TYPE I_B POROSITY FIELD 3

RANGE (6.7, 21.8)



Oil Saturation (S_o)--Field 3

Data Range (13.1, 48.7%) Mean = 33.26%

Class interval = 3% $\sigma = 5.7$

$\alpha_3^2 = .79434377$ $\delta = .072484741$ Type VI_B curve

$r_1 = -4.2828677$ $r_2 = -6.6759047$

$m_1 = 51.961440$ $m_2 = -83.553454$

$C = 37.514620$

$$Y = (37.514620)(t+6.6759047)^{-83.553454}(t+4.2828677)^{51.961440}$$

Curve Range $(-\infty, 57.7)$

VI_B Curve

$\chi^2 = 10.864505$

Fit is good at .50 level

Normal Curve

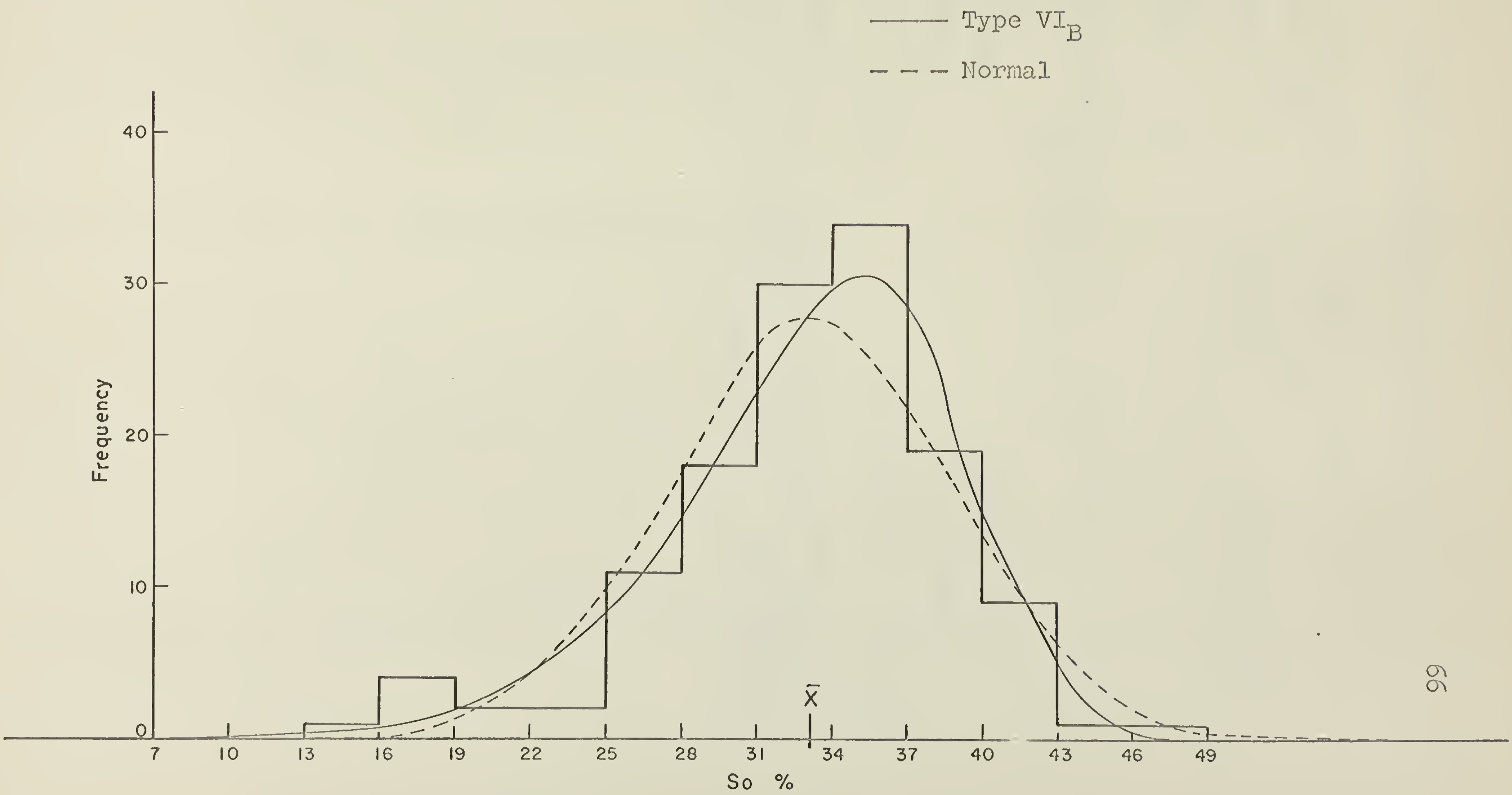
$\chi^2 = 24.761836$

Fit is good at .01 level

S_o Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal (mid-ordinates)
14.5	1	.6	.6	.1
17.5	4	1.3	1.4	.6
20.5	2	2.9	2.9	2.4
23.5	2	5.8	5.9	6.6
26.5	11	11.0	11.1	14.0
29.5	18	18.7	18.7	22.6
32.5	30	26.8	26.6	27.6
35.5	34	30.1	29.7	25.5
38.5	19	23.1	22.8	17.8
41.5	9	9.7	9.9	9.4
44.5	1	1.5	1.8	3.8
47.5	1	.04	.08	1.1
	132			

FIGURE 13 TYPE VI_B OIL SATURATION FIELD 3

RANGE $(-\infty,\ 57.7)$



Water Saturation (S_w)--Field 3

Data Range (30.4, 76.1) Mean = 45.35%

Class interval = 5% σ = 7.85

$\alpha_3^{\,2}$ = 1.4631973 δ = -.03992642 Type III_B curve

A = 1.6534022 C = 13.616174

$$Y = C(A + t)^{A^2-1} \quad e^{-At}$$

$$Y = (13.616174)(1.6534022)^{1.38950} \quad e^{-1.6534022\ t}$$

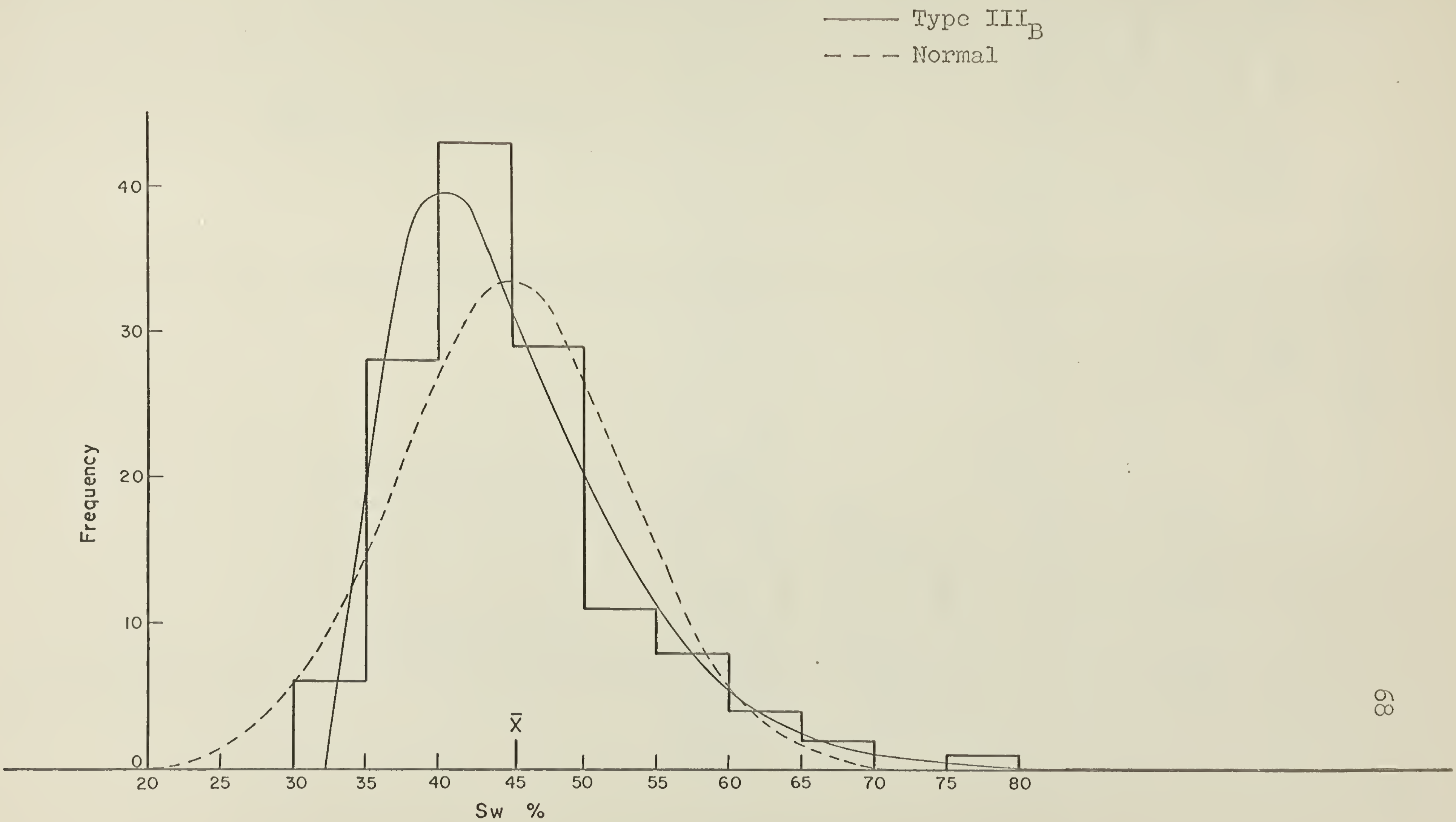
Curve Range (32.4, ∞)

III_B Curve	Normal Curve
χ^2 = 4.7735815	χ^2 = 16.889006
Fit is good at .85 level	Fit is good at .05 level

S_w Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Mid-ordinates
32.5	6	1.1	4.1	9.3
37.5	28	35.1	33.5	21.1
42.5	43	38.2	37.6	31.8
47.5	29	26.3	26.3	32.0
52.5	11	14.9	15.1	21.4
57.5	8	7.6	7.7	9.6
62.5	4	3.6	3.7	2.8
67.5	2	1.6	1.7	.6
72.5	0	.7	.7	.07
77.5	1	.3	.3	.01
	132			

FIGURE 14 TYPE III_B WATER SATURATION FIELD 3

RANGE (32.4, ∞)



Porosity (ϕ)--Field 4

Data Range (9.3, 22.4%) Mean = 17.8%

Class interval = 1.5% σ = 2.0

$\alpha_3^{\,2}$ = 1.1315972 δ = .0047189 Type III_B curve

A = 1.8801133 C = 18.684424

$Y = C(A + t)^{A^2-1} \quad e^{-At}$

$Y = (18.684424)(1.8801133+t)^{2.53440} \quad e^{-1.880113\ t}$

Curve Range (-∞, 21.57)

III_B Curve

χ^2 = 6.3034068

Fit is good at .60

Normal Curve

χ^2 = 19.69985

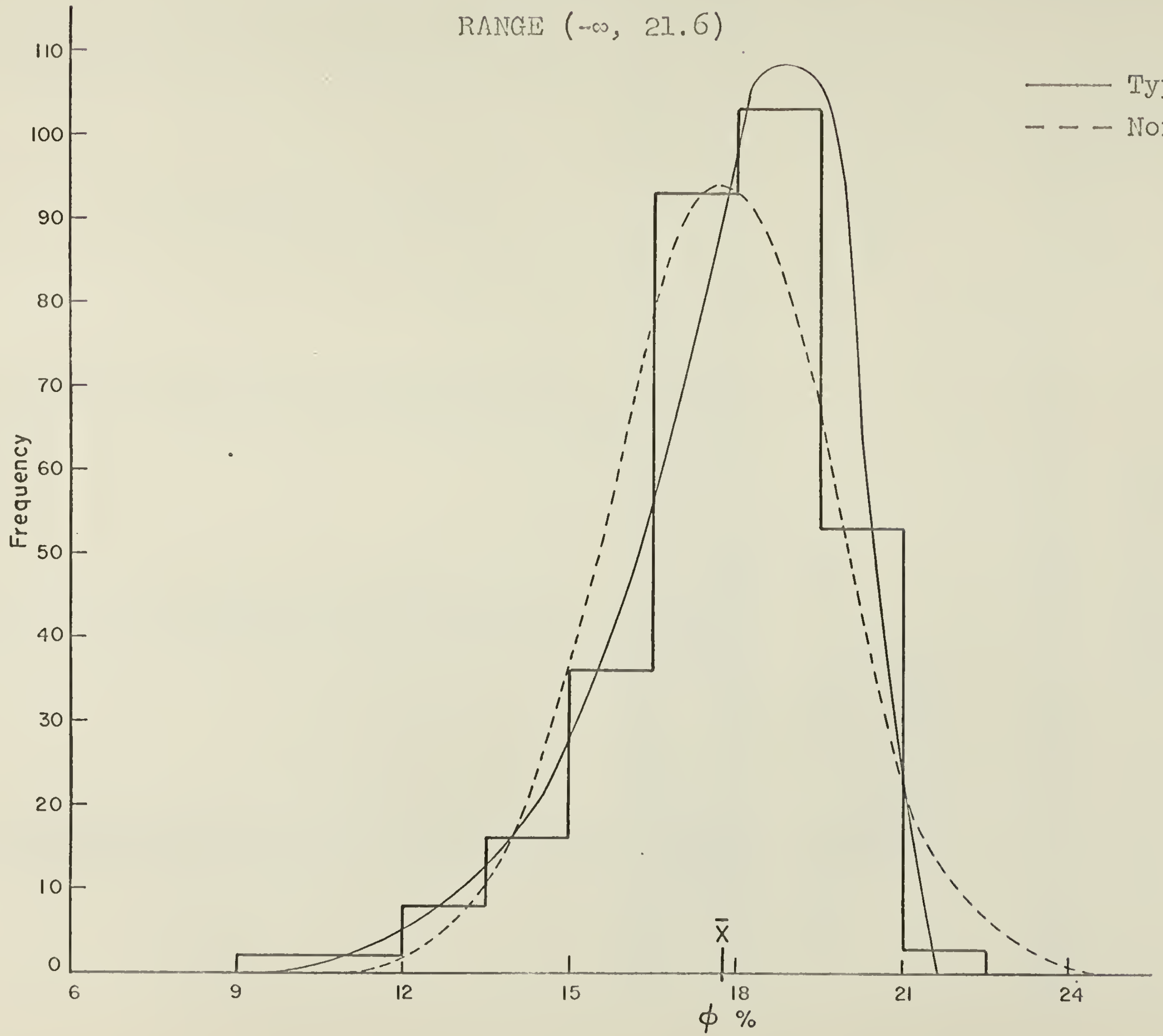
Fit is good at .02

Porosity Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Mid-ordinates
9.75	2	.9	.9	.03
11.25	2	2.6	2.7	.5
12.75	8	7.1	7.3	4.0
14.25	16	18.0	18.6	20.1
15.75	36	41.3	41.9	57.0
17.25	93	79.0	78.8	91.7
18.75	103	108.2	104.9	83.9
20.25	53	61.2	59.3	43.6
21.75	3	0.	2.1	12.9
23.25	0.	0.	0.	2.2
	316			

FIGURE 15 TYPE III_B

RANGE $(-\infty, 21.6)$

POROSITY FIELD 4



Oil Saturation (S_o)--Field 4

Data Range (13.3, 68.2%) Mean = 39.96%

Class interval = 3.5% σ = 10.1

$\alpha_3^{\,2}$ = .01836461 δ = -.13178302 Type I_B curve

r_1 = -3.2859450 r_2 = 4.3142724

m_1 = 4.6968277 m_2 = 6.4796347

$C = 1.1959325 \times 10^{-5}$

$$Y = (1.1959325\text{x}10^{-5})(t+3.285945)^{4.6968277}\ (4.3142724-t)^{6.4796347}$$

Curve Range (6.9, 83.3)

I_B Curve

χ^2 = 13.992946

Fit is good at .60 level

Normal Curve

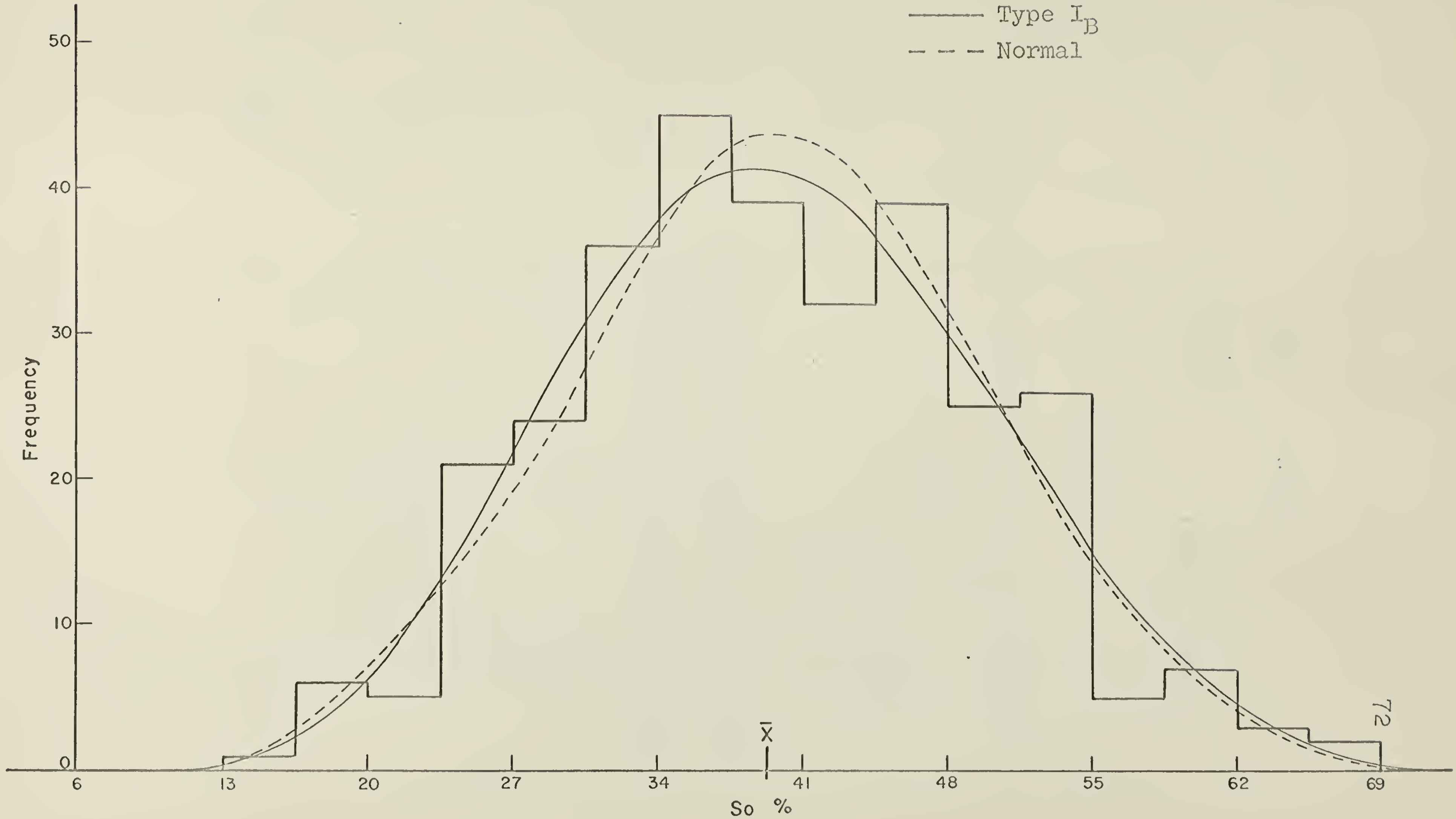
χ^2 = 15.734882

Fit is good at .40 level

S_o Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal (mid-ordinates)
14.75	1	.99	1.1	1.9
18.25	6	3.9	4.0	4.4
21.75	5	9.6	9.7	8.7
25.25	21	17.6	17.7	15.3
28.75	24	26.6	26.6	23.9
32.25	36	34.7	34.5	33.1
35.75	45	40.0	39.9	40.5
39.25	39	41.6	41.5	43.8
42.75	32	39.3	39.2	42.1
46.25	39	33.9	33.8	35.8
49.75	25	26.5	26.5	27.0
53.25	26	18.7	18.7	18.0
56.75	5	11.7	11.8	10.6
60.25	7	6.4	6.5	5.6
63.75	3	3.0	3.0	2.6
67.25	2	1.1	1.1	1.1
	316			

FIGURE 16 TYPE I_B OIL SATURATION FIELD 4

RANGE (6.9, 83.3)



Water Saturation (S_w)--Field 4

Data Range (18.6, 76.8%)   Mean = 39.81%

Class interval = 6% σ = 13.5

$\alpha_3^{\ 2}$ = .44508784 δ = -.41374306   Type I_J curve

r_1 = -1.3112949 r_2 = 2.9237671

m_1 = -.122539 m_2 = .9564568

C = 10.194628

$$Y = (10.194628)(t+1.3112949)^{-.122539}\ (2.9237671-t)^{.9564568}$$

Curve Range (22.1, 79.2)

I_B Curve

χ^2 = 15.406190

Fit is good at .05 level

Normal Curve

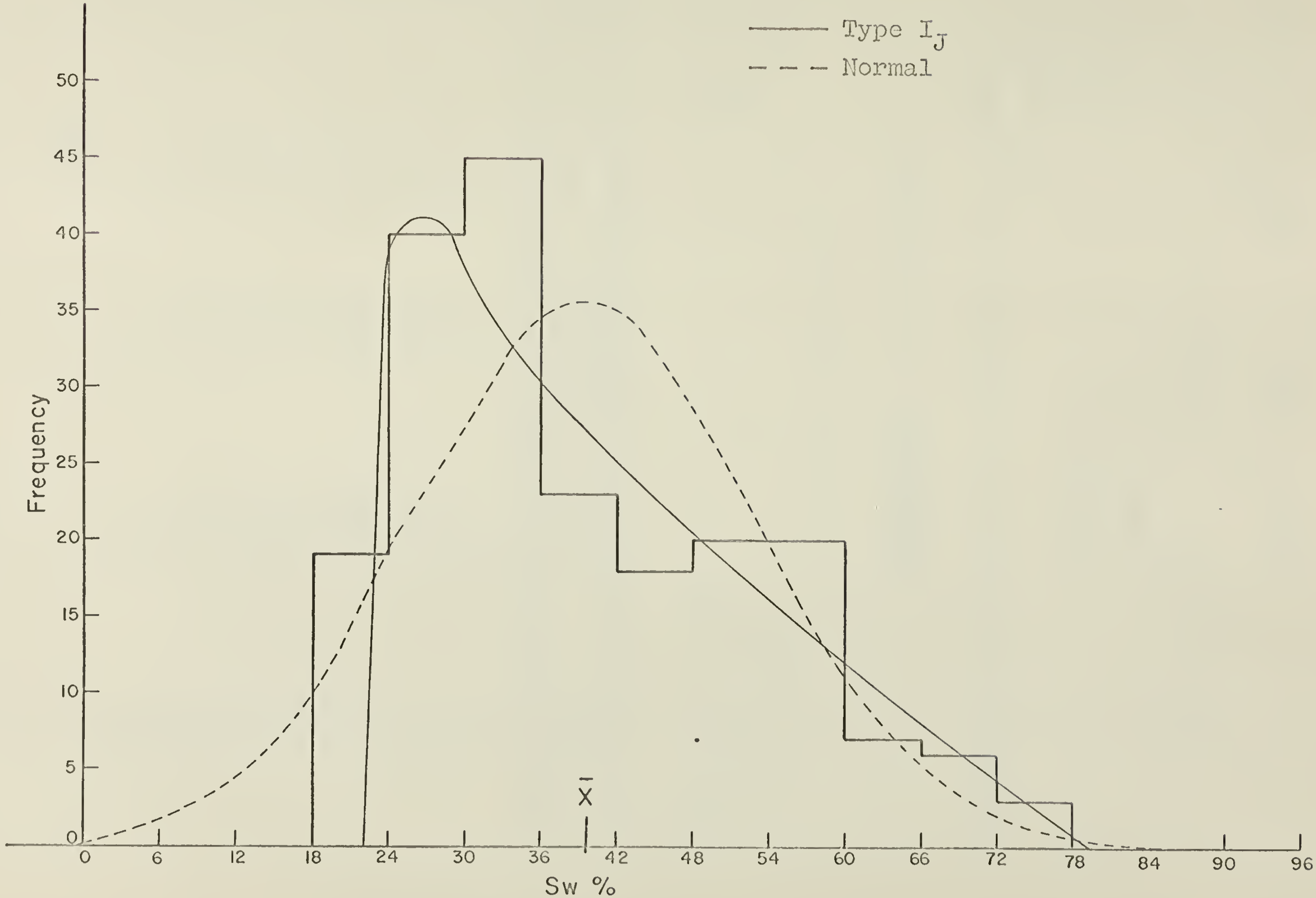
χ^2 = 28.03342

Fit is not good

S_w Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Mid-ordinates
21	19	0	8.0	14.2
27	40	41.3	41.7	23.4
33	45	33.5	33.6	31.8
39	23	27.8	27.9	35.5
45	18	23.0	23.0	32.6
51	20	18.5	18.5	24.6
57	20	14.4	14.4	15.2
63	7	10.4	10.4	7.7
69	6	6.5	6.5	3.2
75	3	2.6	2.6	1.1
	201			

FIGURE 17 TYPE I_J WATER SATURATION FIELD 4

RANGE (22.1, 79.2)



Porosity (ϕ)--Field 5

Data Range (3.4, 34.3%) Mean = 20.66%

Class interval = 3% σ = 4.32

$\alpha_3^{\,2}$ = .33849171 δ = .25960502 Type IV curve

r = 1.1205492 m = 5.8520056

v = -3.9842932 S = 2.7291720

C = 3.3495724×10^7

$$Y = (3.3405724\text{x}10^7)[(t+1.1205492)^2 + (2.7291720)^2]^{-5.8520056} \text{ x}$$

$$e^{3.9842932 \tan^{-1}\left(\frac{t+1.1205492}{2.7291720}\right)}$$

Curve Range ($-\infty$, ∞)

IV Curve

χ^2 = 83.10963

Fit is not good

Normal Curve

χ^2 = 109.93934

Fit is not good

Porosity Midpoint (%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Mid-ordinates
4.5	9	4.3	4.5	4.5
7.5	13	12.2	12.8	4.7
10.5	49	35.5	37.1	30.3
13.5	96	99.5	103.1	120.6
16.5	181	245.4	248.6	296.2
19.5	472	450.6	443.5	449.0
22.5	592	489.4	476.3	420.3
25.5	182	257.0	259.9	242.9
28.5	49	65.0	71.1	86.6
31.5	29	10.0	11.7	19.1
34.5	1	1.3	1.5	2.6
	1673			

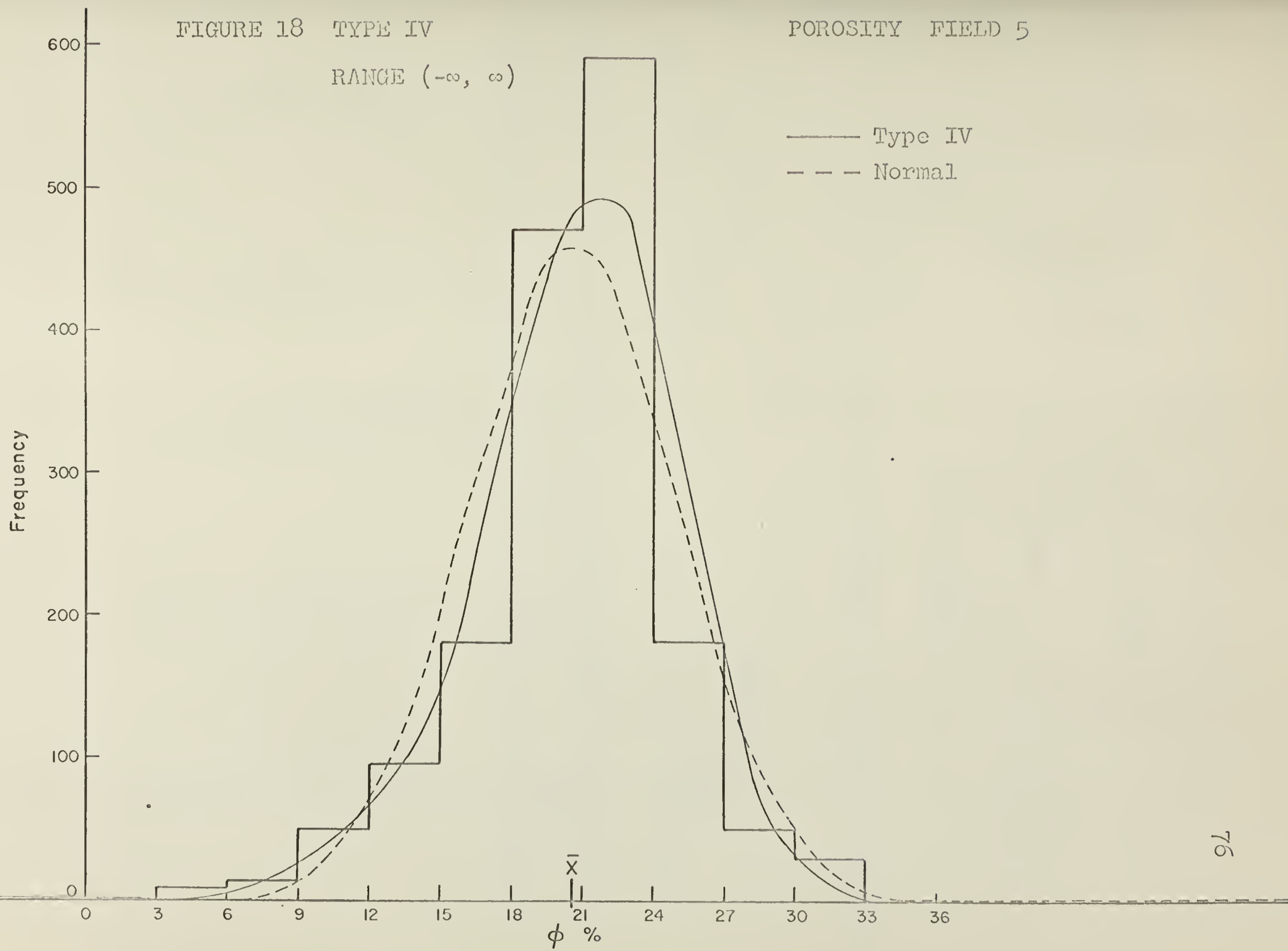


FIGURE 18 TYPE IV

RANGE $(-\infty, \infty)$

POROSITY FIELD 5

Oil Saturation (S_o)--Field 5

Data Range (5.7, 77.2%) Mean = 32.77%

Class interval = 8% σ = 13.84

α_3^2 = .00047316776 δ = -.11483628 Type I_B curve

r_1 = -3.9580728 r_2 = 4.1474939

m_1 = 6.5279175 m_2 = 6.8881809

$C = 2.5711286 \times 10^{-6}$

$$Y = (2.5711286\text{x}10^{-6})(t+3.9580728)^{6.5279175}\ (4.1474939-t)^{6.8881809}$$

Curve Range (-21.9, 90.1)

I_B Curve

χ^2 = 16.2566686

Fit is good at .05 level

Normal Curve

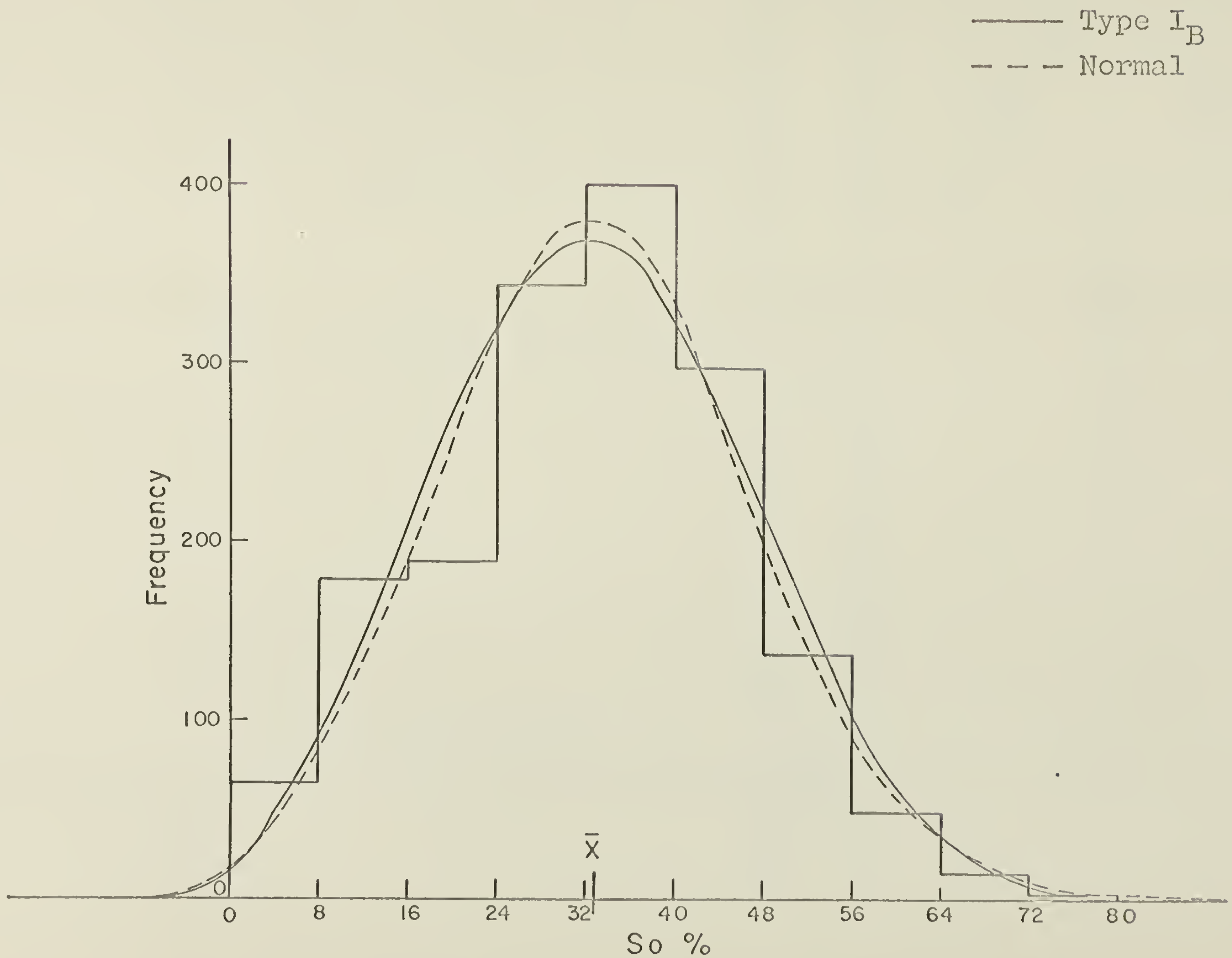
χ^2 = 17.879294

Fit is good at .04 level

S_o Midpoint(%)	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Mid-ordinates
4	65	48.6	50.9	46.1
12	178	141.0	142.4	128.5
20	189	263.4	262.0	256.4
28	343	353.6	349.9	366.0
36	399	357.5	353.6	373.8
44	297	273.6	272.1	273.1
52	136	153.7	154.8	142.8
60	48	58.4	60.6	53.4
68	14	12.5	13.9	14.3
76	3	.9	1.3	2.7
	1672			

FIGURE 19 TYPE I_B OIL SATURATION FIELD 5

RANGE (-21.9, 90.1)



Water Saturation (S_w)--Field 5

Data Range (12.4, 92.8%) Mean 52.17

Class interval 8% $\sigma = 16.0\%$

$\alpha_3^{\,2} = .000611895$ $\delta = -.02569028$ Normal Type curve

C = 244.82

$$Y = 244.82 \; e^{-\frac{t^2}{2}}$$

Curve Range $(-\infty, \infty)$

Normal

$\chi^2 = 14.359695$

Fit is good at .20 level

S_w Midpoint (%)	Frequency	Normal Mid-ordinates
12	15	9.7
20	40	30.2
28	56	72.7
36	111	136.3
44	237	198.8
52	242	225.6
60	189	199.2
68	106	136.9
76	76	73.2
84	47	30.4
92	10	9.9
	1129	

Water Saturation (S_w)--Field 5

Data Range (12.4, 92.8%) Mean 52.17

Class interval 8% σ = 16.0%

$\alpha_3{}^2$ = .000611895 δ = -.02569028 Normal Type curve

C = 244.82

$$Y = 244.82\ e^{-\frac{t^2}{2}}$$

Curve Range $(-\infty, \infty)$

Normal

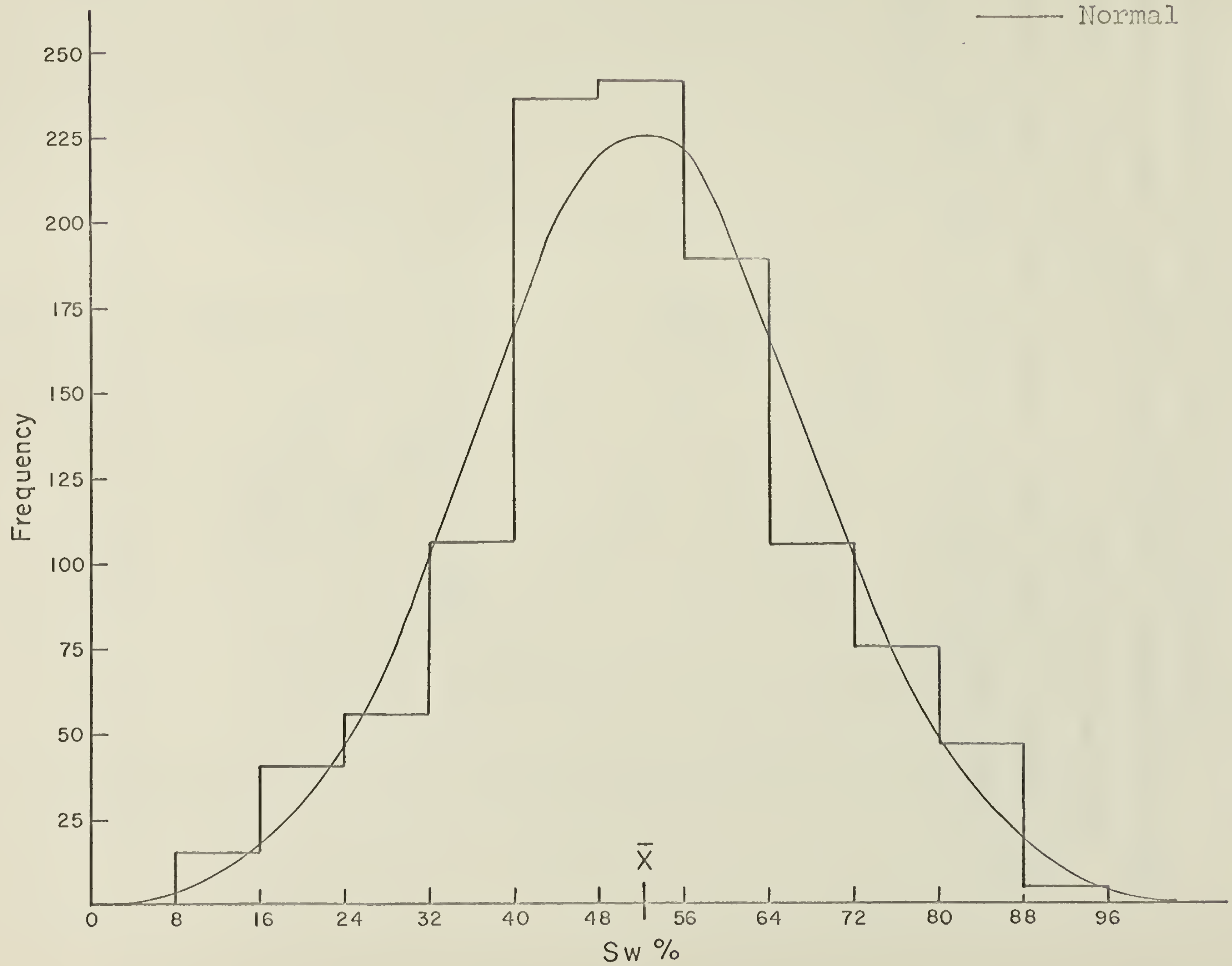
χ^2 = 14.359695

Fit is good at .20 level

S_w Midpoint (%)	Frequency	Normal Mid-ordinates
12	15	9.7
20	40	30.2
28	56	72.7
36	111	136.3
44	237	198.8
52	242	225.6
60	189	199.2
68	106	136.9
76	76	73.2
84	47	30.4
92	10	9.9
	1129	

FIGURE 20 NORMAL WATER SATURATION FIELD 5

RANGE $(-\infty, \infty)$



<u>Summary</u>

A concise and easily visualized summary of the distributions of the data may be shown by means of the (α_3^2, δ) chart. Figure 2 is the complete (α_3^2, δ) chart; however, for purposes of showing the distributions of the properties under consideration, only the necessary portion of the chart has been prepared for each of the field properties.

Figure 21 shows the distributions for porosity for the five fields. There appears to be no pattern to the types of distributions, in that Fields 1 and 3 are of the Type I_B, Field 4 is a Type III_B, Field 2 is a Type VI_B and Field 5 is a Type IV.

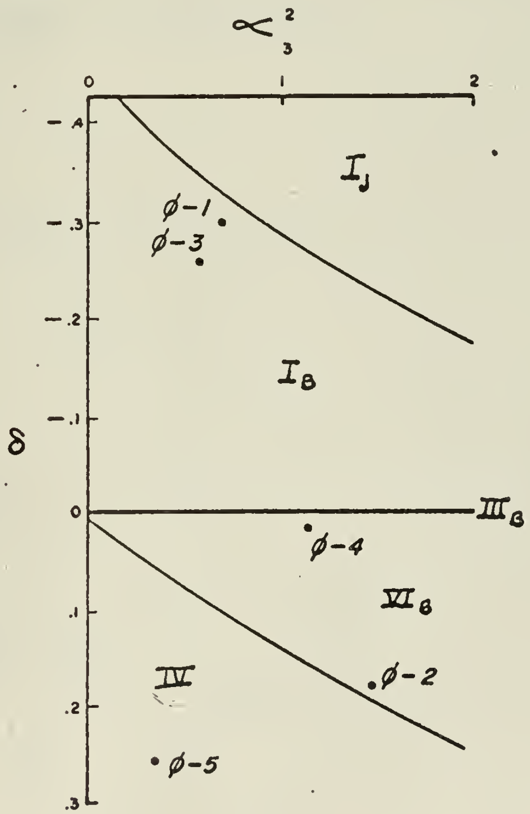


FIGURE 21 $(\alpha_3^2; \delta)$ Chart of Porosity (Normal Curve is Pt. 0.0)

Of the five fields, only Field 4 satisfied the goodness of fit test for the normal curve. This is somewhat in contrast to the conclusion of Jan Law[11] that:

> "With some exceptions permeability and porosity assemblages give respectively satisfactory logarithmic and arithmetic normal frequency distributions."

For the present study it was the exception where porosity data gave a satisfactory arithmetic normal frequency distribution.

The (α_3^2, δ) chart for the S_o data, Figure 22, indicates a somewhat different situation regarding normal distributions. Four of the five fields, 1, 3, 4, 5, satisfied the goodness of fit test for normal distributions.

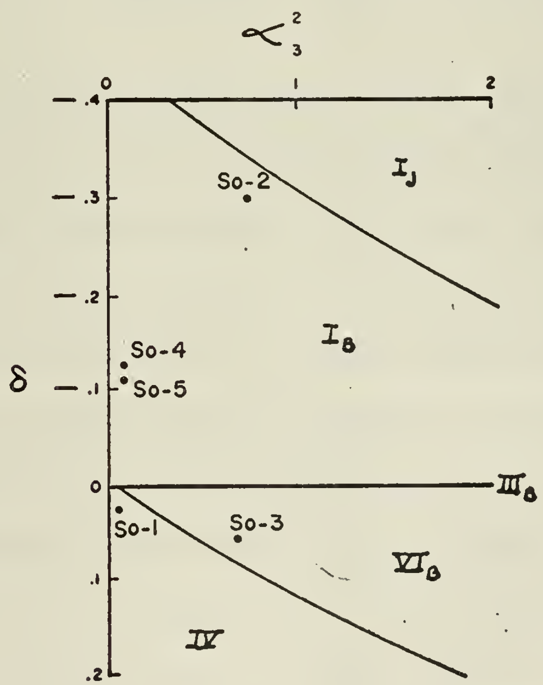


FIGURE 22 (α_3^2, δ) Chart of S_o

All five fields for S_w were fitted by a Type I curve with varying values of α_3^2 and δ. Field 5 satisfied the goodness of fit test for the normal curve.

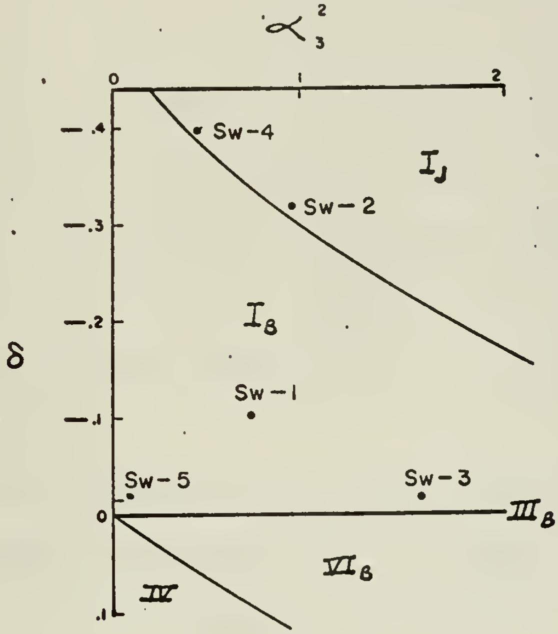


FIGURE 23 (α_3^2, δ) Chart of S_w

It is noted that both the S_o and S_w distributions for Field 5 which has a considerably greater number of samples included, approximately 1700, than any of the other fields satisfies the goodness of fit for the normal curve. This could indicate that if large enough samples are taken, the S_o and S_w distributions approach normality. For sampling considerations, the assumption that S_o and S_w populations are normally distributed would permit many established techniques to be applied to the analysis of these properties.

A study was made of what effect the choice of interval size would have on the α_3^2 and δ values. For each set of data various interval widths were used, i.e., for porosity

1 percent, 1.5 percent, 2 percent, and 3 percent interval widths, to calculate the parameters α_3^2 and δ. For all fields with the exception of Field 3, no appreciable difference was noted in the values of α_3^2 and δ for the various interval sizes tried. There were only 132 observations for Field 3, and the calculated values of α_3^2 and δ were more sensitive to the size of interval used, with α_3^2 ranging from .641 to .279 and δ ranging from -.249 to -.099 as the interval size was varied from 1 to 2.5 percent in steps of .5 percent.

It was concluded from this study that with a sufficiently large number of observations (> 200), a width chosen to give between 7 to 15 intervals will give consistent results in the calculation of α_3^2 and δ.

Although it was not possible to study enough fields to determine the relation between reservoir type and the parameters α_3^2 and δ herein measured, it would be reasonable to assume that such a situation exists. Specifically the parameters α_3^2 and δ would be expected to reflect the variation of individual properties throughout a single depositional unit. They should also characterize the depositional unit, distinguishing it from similar units in a geologic basin or province. Within this context, the variables α_3^2 and δ could represent properties which could be correlated in the geologic sense for determining stratigraphic equality.

If data were available in sufficient quantity to yield α_3^2 and δ parameters for individual wells, then these

variables could be contoured for specific reservoirs or used to segment the reservoir into smaller more homogeneous units for mathematical analysis. The values themselves would then provide the information necessary for preparation of a mathematical model of the reservoir including the effects of heterogeneity.

Thus in reservoir models there is no "a priori" reason why the Gaussian normal distribution function need be used to study the reservoir performance. If experimental evidence indicates the existence of a non-normal Pearson type distribution function for a particular property then this specific distribution function could be used.

However before quantitative use is made of the statistics α_3^2 and δ, it should be emphasized that these apply only to the set of data from which they were extracted. They are only estimates of the parameters α_{3r}^2 and δ_r where the subscript r denotes the value of parameters for the population of samples which comprise the entire reservoir. For the normal distribution which is described by the mean and standard deviation, techniques are well known that employ these sample statistics, $\overline{X}$ and σ, to estimate population parameters. However very little information is available for estimating population parameters α_{3r}^2 and δ_r from the sample statistics α_3^2 and δ.

CHAPTER IV

ANALYSIS OF LOGRITHMS OF DATA

Introduction

Jan Law[1] in his work with core data concluded that "with some exceptions permeability....assemblages give satisfactory logrithmic....normal frequency distributions." To investigate the applicability of Jan Law's conclusion to the core data available for the fields under study is the purpose of this chapter.

It can be observed from Figures 4 and 8 that the frequency distributions of the permeability of Fields 1 and 2 are markedly skewed. The distributions are of the Pearson Type I_J. Does the distribution of the logrithms of these data approximate the normal curve?

Logrithmic Distributions

If the distributions of these data do satisfy a logrithmic normal frequency distribution, an accumulated frequency curve of the permeability data plotted on logrithmic probability paper should follow a straight line.[2]

The sample data may be cumulated and put in percentage form as in Tables V and VI. These cumulative percentages may then be plotted on logrithmic probability paper. If the resulting curve is approximately a straight line, we may proceed with assurance to fit a normal curve to the logrithms of the data.

Three cycle logrithmic probability paper was used to plot the cumulative percentages from Tables V and VI.

TABLE V Cumulative Distribution of Permeability for Field 1

Permeability in md.	Number of Measurements	Percent of Total
less than 1.60	130	10
2.89	195	15
6.25	260	20
11.00	325	25
16.00	390	30
24.00	455	35
30.00	520	40
37.00	585	45
45.00	651	50
53.00	715	55
65.00	780	60
71.00	845	65
84.00	910	70
97.00	975	75
111.00	1040	80
126.50	1105	85
151.50	1170	90
185.00	1235	95
439.00	1303	100

TABLE VI Cumulative Distribution of Permeability for Field 2

Permeability in md.	Number of Measurements	Percent of Total
less than 1.1	31	5
1.8	63	10
3.7	94	15
7.0	126	20
10.0	158	25
15.0	189	30
20.0	220	35
24.9	252	40
29.0	283	45
32.0	317	50
36.0	346	55
42.0	378	60
48.0	409	65
53.0	441	70
63.0	472	75
73.0	504	80
90.0	535	85
122.8	572	90
175.0	603	95
380.0	635	100

The resultant curves obtained for the two fields are shown in Figures 24 and 25 respectively. From visual observation it appears that the two curves do not approximate straight lines. Therefore it may be concluded that the permeability data for these fields do not follow satisfactorily the logrithmic normal distribution.

The question that now arises is if the logrithms of the data are not normally distributed, what distribution pattern do they follow? To answer this question, the permeability data were converted to logrithms.

The logrithms of the permeability of the two fields were then treated as data to which a Pearson frequency curve was fitted. The calculated results using the logrithms are shown on the following pages, with the calculations and figures for Fields 1 and 2 respectively shown.

Fitting the Logrithmic Data to Pearson Curves

The procedure outlined in Chapter III was used to fit the logrithms of the permeability of Fields 1 and 2 to a Pearson type curve. The curves obtained are shown in Figures 26 and 27 and were found to be Type I_J and I_B. Additionally a normal curve fitted to the histogram is shown for comparative purposes.

To simplify the interpretation of Figures 26 and 27, the abscissa of the curves were converted to an interval scale which is denoted as a π scale. Tables VII and VIII show the logrithm value and the millidarcy value of each interval of the abscissa.

FIGURE 24 PERMEABILITY MEASUREMENTS--FIELD 1 (On Logrithmic Probability Paper)

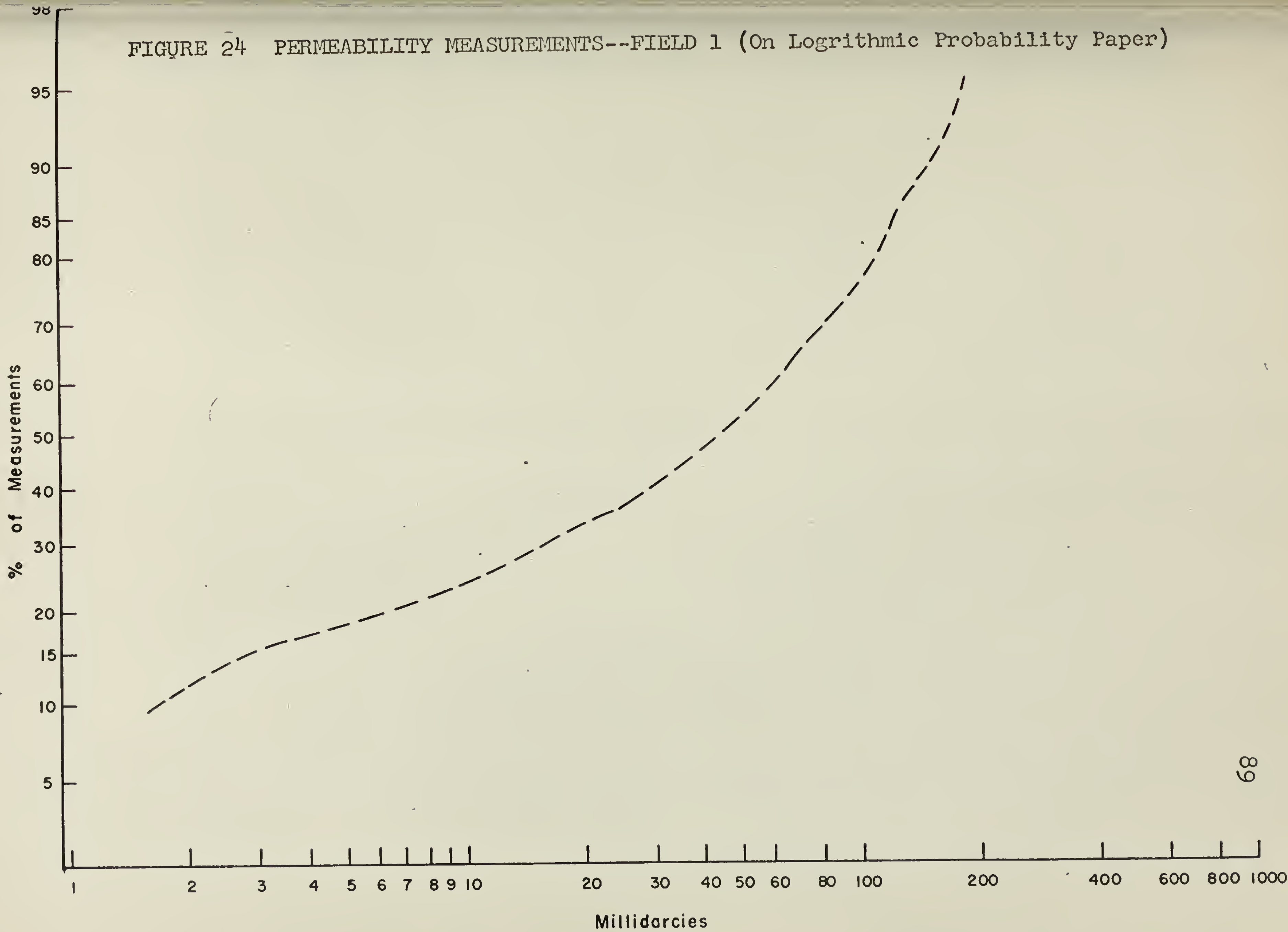


FIGURE 25 PERMEABILITY MEASUREMENTS--FIELD 2 (On Logrithmic Probability Paper)

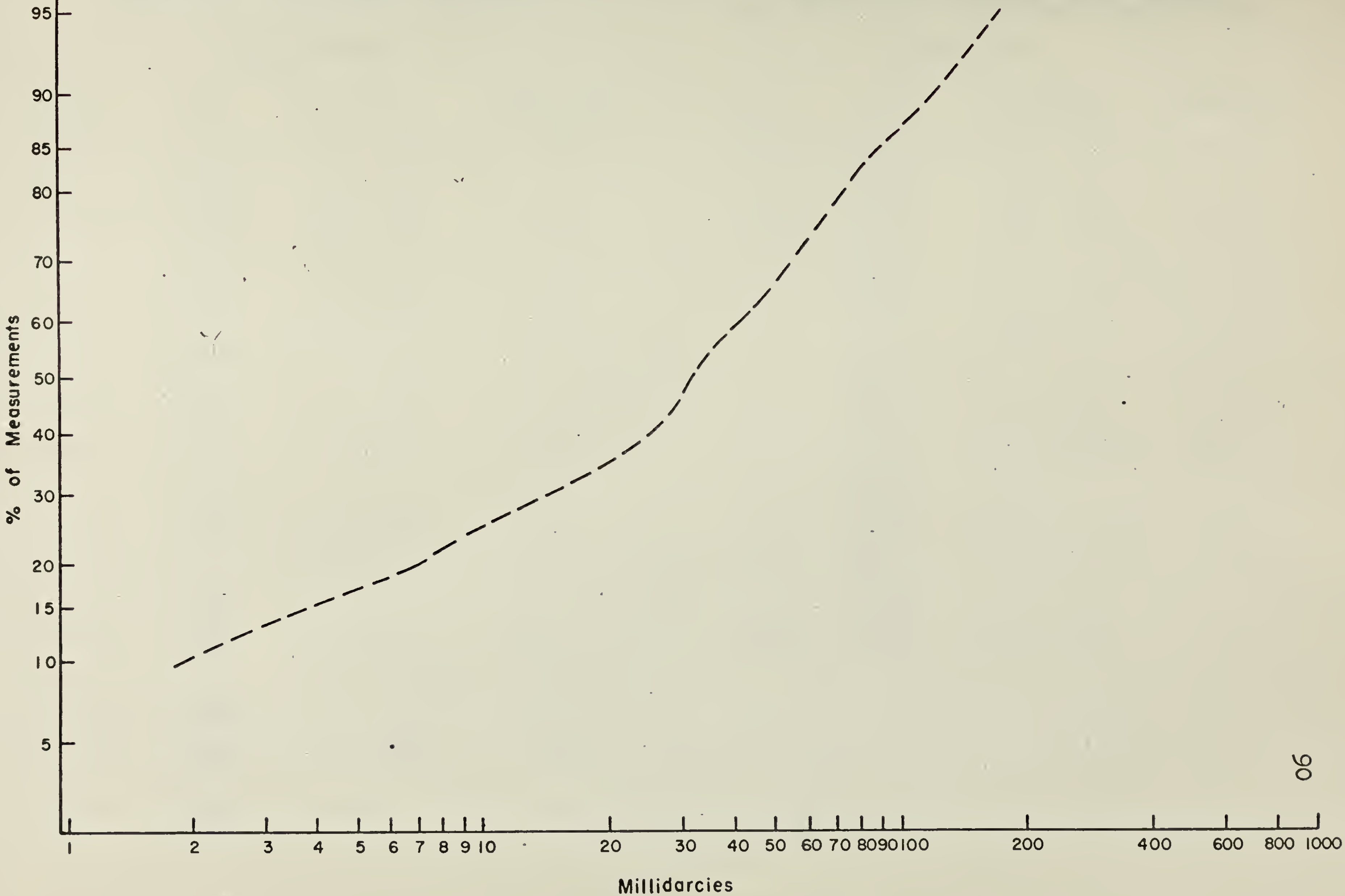


TABLE VII Logrithmic Frequency Distribution of Field 1 (Permeability)

π Scale	Log K (10^2)	Class Limits Millidarcy	Frequency
1- 2	1.3001-1.6000	.134-.398	17
2- 3	1.6001-1.9000	.399-.784	53
3- 4	1.9001-2.2000	.785-1.584	59
4- 5	2.2001-2.5000	1.585-3.16	74
5- 6	2.5001-2.8000	3.17-6.25	57
6- 7	2.8001-3.1000	6.26-12.6	89
7- 8	3.1001-3.4000	12.61-25.5	132
8- 9	3.4001-3.7000	25.6-50.1	209
9-10	3.7001-4.0000	50.2-100.	297
10-11	4.0001-4.3000	100.1-199.4	266
11-12	4.3001-4.6000	199.5-398	49
12-13	4.6001-4.9000	399-784	1

TABLE VIII Logrithmic Frequency Distribution of Field 2 (Permeability)

π Scale	Log K (10^2)	Class Limits Millidarcy	Frequency
1- 2	1.201-1.600	.159-.398	9
2- 3	1.601-2.000	.399-1.000	20
3- 4	2.001-2.400	1.001-2.51	52
4- 5	2.401-2.800	2.52-6.31	37
5- 6	2.801-3.200	6.32-15.9	76
6- 7	3.201-3.600	15.91-39.80	163
7- 8	3.601-4.000	39.81-100.00	178
8- 9	4.001-4.400	100.01-251.0	90
9-10	4.401-4.801	251.01-631.	5

Calculations log K--Field 1

Data Range (-1.30103, 2.64246) $Mean_{log}$ = 1.4288151

Class Interval = .30000 σ = .74250

$\alpha_3{}^2$ = .86963292 δ = -.47453376 Type I_J Curve

r_1 = -1.0619523 r_2 = 3.0271256

m_1 = -.42484190 m_2 = .63950490

C = 7.5824862

$$Y = 7.5824862(t+1.0619523)^{-.42484190}(3.0271256-t)^{.63950490}$$

$Range_{log}$ (-1.17381, 2.21681)

I_J Curve

χ^2 = 30.534208

Fit is good at .001 level

Normal Curve

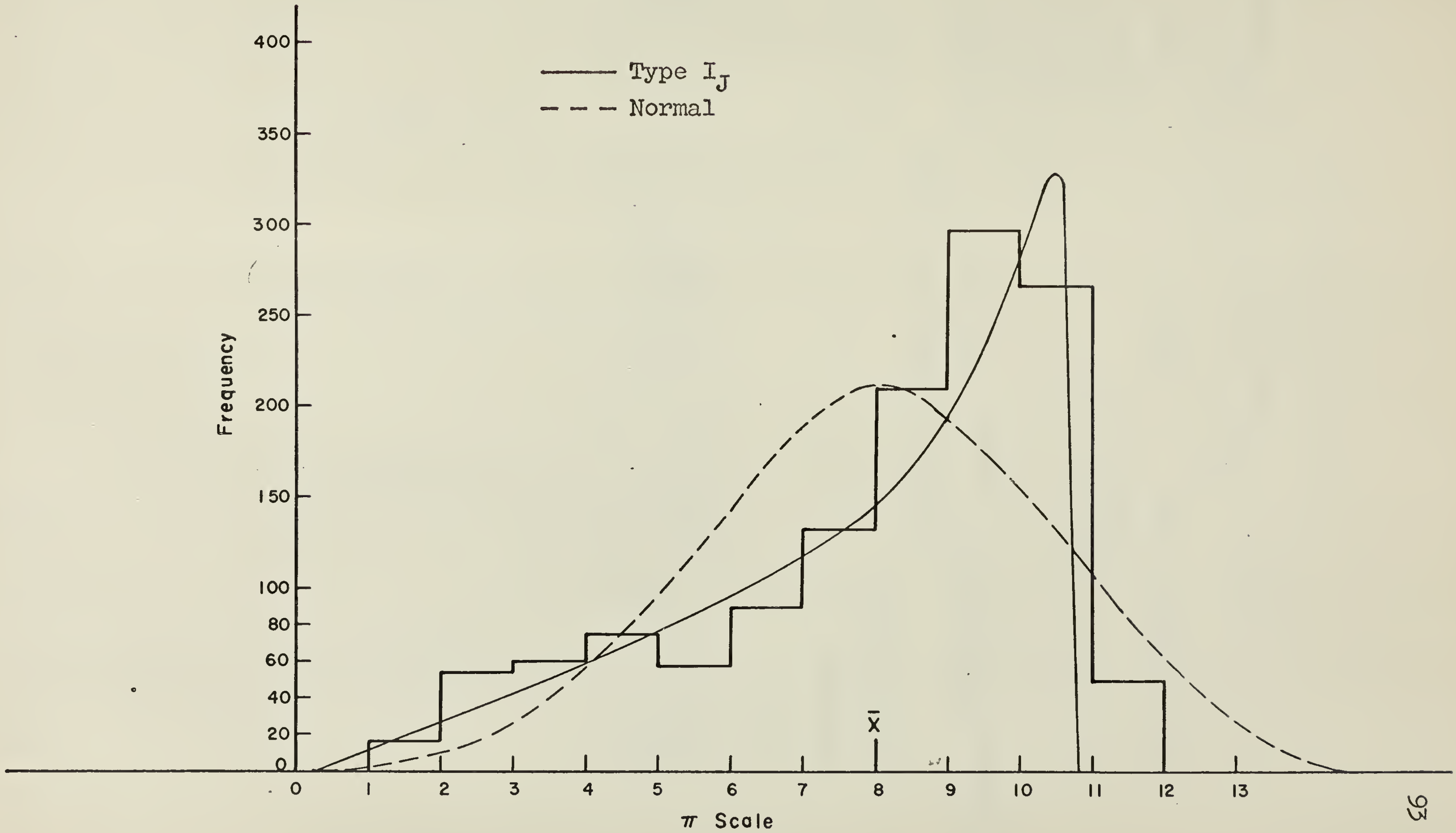
χ^2 = 428.38022

Fit is not good

π Scale Midpoint	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Curve Mid-ordinates
1.5	17	22.8	22.6	6.1
2.5	53	38.5	38.5	16.4
3.5	59	53.3	53.3	37.6
4.5	74	68.6	68.7	73.3
5.5	57	85.5	85.6	121.3
6.5	89	105.4	105.6	170.5
7.5	132	130.6	131.0	203.8
8.5	209	166.7	167.8	206.9
9.5	297	231.7	236.6	178.6
10.5	266	308.8	289.1	130.9
11.5	49	0.	0.0	81.6
12.5	1	0.	0.0	43.2

FIGURE 26 FREQUENCY DISTRIBUTION OF LOG K

FIELD 1 TYPE I_J



Calculations log K--Field 2

Data $\text{Range}_{\log}$ (-1.2040, 2.5800) $\text{Mean}_{\log} = 1.373706$

Class Intervals = .4000 $\sigma = .6920$

$\alpha_3{}^2 = .625648$ $\delta = -.2240993$ Type I_B Curve

$r_1 = 1.5577193$ $r_2 = 5.0873113$

$m_1 = .62325930$ $m_2 = 4.3013559$

$C = 8.9991354 \times 10^{-2}$

$$Y = (8.9991354\text{x}10^{-2})(t+1.5577193)^{.62325930}(5.0873713-t)^{4.3013559}$$

$\text{Range}_{\log}$ (-2.14621, 2.473790)

I_B Curve

$\chi^2 = 24.989362$

Fit is good at .005 level

Normal Curve

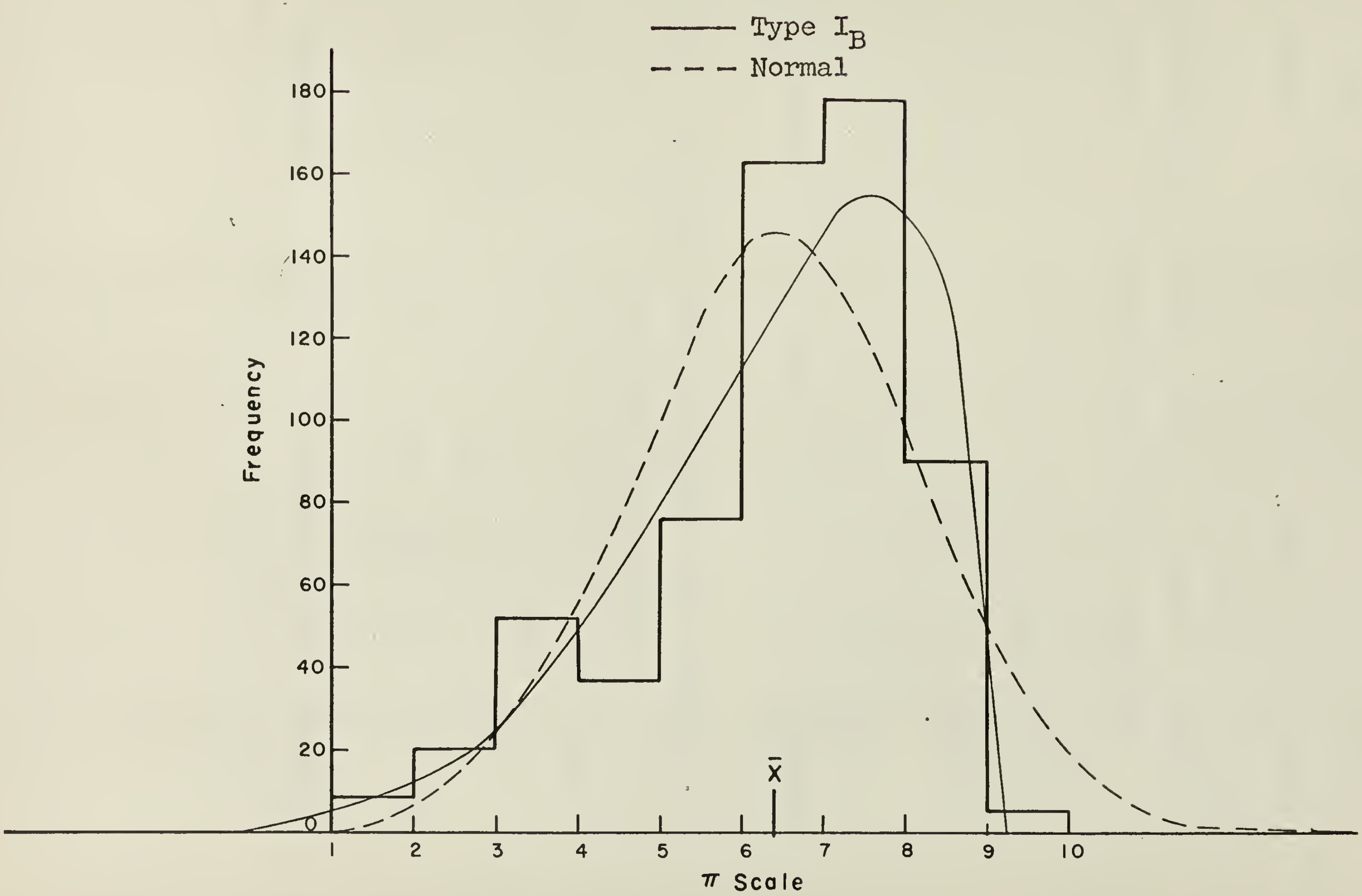
$\chi^2 = 108.47524$

Fit is not good

π Scale Mid-point	Frequency	Graduation Mid-ordinates	Graduation Areas	Normal Curve (mid-ordinates)
1.5	9	7.1	7.4	2.5
2.5	20	17.7	18.0	10.9
3.5	52	35.7	36.1	34.5
4.5	37	62.3	62.6	78.1
5.5	76	96.1	96.2	126.4
6.5	163	131.7	131.2	146.5
7.5	178	155.0	153.1	121.4
8.5	90	130.1	122.1	72.1
9.5	5	0.0	9.9	30.6

FIGURE 27 FREQUENCY DISTRIBUTION OF LOG K

FIELD 2 TYPE I_B



Summary

From the cumulative frequency curves and the logrithmic frequency distribution curves it may be concluded that the permeability data under study does not satisfactorily fit a logrithmic normal curve. It is not the intent of this study to dispute Jan Law's conclusions that the logrithms of permeability data approximate a normal curve, but more to express a word of caution in indiscrimately using the logrithmic normal curve to analyze the permeability distribution of a particular set of data. The data used by Jan Law came from a different geographical area and from fields with different depositional properties than the data for this study. Therefore it appears that each and every set of data must be analyzed to determine its applicability to the logrithmic normal curve.

Logrithms do provide a means to remove some skewness from a set of data, but not necessarily to convert it to a normal distribution. If sampling techniques were developed for Pearson's Type I_B distributions, the conversion of permeability data to logrithms for analysis could serve a very useful purpose. At present though, this study indicates that caution must be exercised in placing too great of a degree of reliability on conclusions drawn from the analysis of the distributions of the logrithms of permeability data.

CHAPTER V
ANALYSIS OF VARIANCE

Introduction

From each core analysis there may be readily obtained a sample mean ($\overline{X}$) and a sample variance (S_X^2) where:

$$\overline{X}_{well} = \frac{\Sigma X_w}{n} \tag{5-1}$$

and,

$$S_X^2 = \frac{\Sigma(X_w - \overline{X})^2}{n} \tag{5-2}$$

where X_w = individual measurements within each well sample

n = number of samples.

Within a field the various means obtained from the different wells most likely will not be numerically identical. It is desired to determine whether or not the difference between the various means, $\overline{X}_1$, $\overline{X}_2$ .... $\overline{X}_n$ can be explained by random errors, that is, the means are statistically the same and represent identical populations, or whether the means are actually different where the differences do not result from sampling errors.

To illustrate the techniques involved in making these determinations, the data obtained from the core analysis for porosity of Field 2 will be used. From this data the following sample means and sample variances were calculated for the wells cored:

Well	Number of Measurements	Mean ($\overline{X}_\phi$)	Variance ($S_{X_\phi}^{\ 2}$)
1	24	.199	.000198
2	28	.189	.000202
3	26	.200	.000128
4	27	.205	.000255
5	18	.199	.000263
6	36	.186	.000890
7	14	.184	.000013
8	23	.208	.000174
9	34	.191	.000607
10	20	.190	.000228
11	31	.203	.000402
12	19	.208	.000894
13	15	.181	.000946
14	30	.212	.000312
	345	2.775	

From the above it can be observed that the means range from .181 to .212. The question which arises, is whether ordinary random sampling errors account for the differences in these means, or may it be concluded that the means are different because of reasons other than sampling fluctuations?

The F Test

The F test devised by R. A. Fisher[1] and named in his honor is a means for answering the above question. The basis for this test lies in the availability of two independent estimates of the population variance. Consider a single core analysis W1ϕ consisting of all of the porosity measurements from Well 1. The variance $S_{W1\phi}^2$ of that set is a measure of the internal scatter of the measurements of the

population. Similar statements apply to $S^2_{W2\phi}$, $S^2_{W3\phi}$ .... $S^2_{Wn\phi}$. However, these variances are influenced by sampling error and are not necessarily equal to the population variance. A better estimate of the internal scatter of the measurements may be obtained by pooling the individual-sample variances. The better estimate of the within-sample or error variance is given by:[2]

$$(5\text{-}3)\quad Se^2 = \frac{\Sigma^{n_1}(X_{W1}-\bar{X}_{W1})^2 + \Sigma^{n_2}(X_{W2}-\bar{X}_{W2})^2 + \ldots.\ \Sigma^{n_k}(X_{Wk}-\bar{X}_{Wk})^2}{(n_1 + n_2 \ldots. + n_k) - n_i}$$

where n_i denotes the number of wells and n_1, n_2 .... n_k denotes the number of measurements in wells W1, W2 .... Wk.

The numerator of equation (5-3) is called the "within-groups sum of the squares", and Se^2 may be called the mean square of individual measurements.

The degrees of freedom associated with Se^2 are:[2]

$$(5\text{-}4)\qquad fe = \Sigma^{n_i}\ (nk)_i - n_i$$

where n_i denotes the number of wells and $(nk)_i$ denotes the number of measurements in the ith well, or fe = the total number of measurements in all wells less the number of wells.

Another independent estimate of the population variance may be obtained from the sample means. The variance of the population of means is:[3]

$$(5\text{-}5)\qquad S^2_{m,p} = \frac{\Sigma^{n_i}\ (\bar{X}_i-\bar{X}_p)^2}{n_i - 1}$$

where:

(5-6)
$$\overline{X}_p = \frac{\Sigma \overline{X}_{W1} + \overline{X}_{W2} + \ldots . \overline{X}_{Wk}}{\text{Number of Wells}}$$

However, the variance of the population of means is not in itself equal to the population variance. If all samples were of the same size, the second estimate of the population variance would be:[3]

(5-7)
$$S_p^{\ 2} = n_k \ S^2_{\ m,p}$$

If the samples are of different size, $n_1 \neq n_2 \neq n_k$, which is usually the case in core analysis, a pooled result is to be used as an average sample size,[3]

(5-8)
$$n_o = \frac{1}{n_i - 1} \left(\Sigma^{n_o} \, n_i (n_k)_i - \frac{\Sigma^{n_i} [(n_k)_i]^2}{\Sigma^{n_i} \, (n_k)_i} \right.$$

Therefore:[3]

(5-9)
$$S_p^{\ 2} = n_o \ S^2 \ m,p$$

when $S_p^{\ 2}$ is estimated by either equation (5-7) or (5-9), n_i-1 degrees of freedom are involved. $S_p^{\ 2}$ is often referred to as the mean square of sample means.

If the random factors which give rise to the within-sample or error variance Se^2 are the only factors causing the differences between the sample means, the two independent variance estimates Se^2 and $S_p^{\ 2}$ should, except for sampling error, be equal. The probability of particular rations of $S_p^{\ 2}$ to Se^2 has been computed by Fisher. This distribution

is the well known F distribution[4] which is a function of the degrees of freedom associated with both ${S_p}^2$ and Se^2. Therefore the ratio, F,[3]

(5-10) $$F = \frac{{S_p}^2}{Se^2}$$

is a measure of whether or not random sampling error can account for the observed differences between sample means. By the use of the F table and the degrees of freedom associated with ${S_p}^2$ and Se^2, a determination of whether random sampling errors could account for the differences between the means can be made.

Conducting the F test on the porosity core analysis results for Field 2 gives:

(5-3) $$Se^2 = \frac{\Sigma^{n_1}(X_{W1}-.199)^2 + \Sigma^{n_2}(X_{W2}-.189)^2 + \ldots\ldots}{(24 + 28 \ldots\ldots) - 14}$$

$$Se^2 = .0004559$$

(5-5) $$S^2_{m,p} = \frac{\Sigma^{n_i}(\overline{X}_i - \overline{X}_p)^2}{n_i - 1}$$

(5-6) $$\overline{X}_p = \frac{.199 + .189 + .200 \ldots\ldots + .212}{14}$$

$$\overline{X}_p = \frac{2.755}{14} = .197$$

(5-5) $$S^2_{m,p} = \frac{(.199-.197)^2 + (.189-.197)^2 \ldots\ldots (.212-.187)^2}{14 - 1}$$

$$S^2_{m,p} = 9.128 \times 10^{-5}$$

(5-9) $$S_p^{\,2} = n_o\, S^2_{m,p}$$

(5-8) $$n_o = \frac{1}{14-1}\left(345 - \frac{\Sigma\ 24^2 + 28^2 + 26^2 \ldots\ldots + 30^2}{\Sigma\ 24 + 28 + 26 \ldots\ldots + 30}\right.$$

$$n_o = 24.507$$

Therefore:

$$S_p^{\,2} = (24.507)(9.128 \times 10^{-5}) = .002237$$

(5-10) $$F = \frac{S_p^{\,2}}{Se^2} = \frac{.002237}{.000456}$$

$$F = 4.907$$

Since $S_p^{\,2}$ is associated with 14 - 1 = 13 degrees of freedom and Se^2 with 345 - 14 = 331 degrees of freedom, the F table[4] gives F = 1.76 at the 5 percent confidence level,

F = 2.14 at the 1 percent confidence level.

The F value of 2.14 at the 1 percent level indicates that there is only 1 chance out of 100 that the observed differences between means can be explained by random errors if the calculated F exceeds 2.14. In this case 4.907 > 2.14, therefore it is reasonable to conclude that the populations represented by the sample means are actually different and the difference cannot be explained by random variation.

Since there is an actual difference between the 14 sample means, further testing must now be conducted to attempt to determine if there are sub-groupings of these means that are statistically homogeneous. For this field,

the wells were lumped together by leases and an F test conducted to determine if the means of the three leases were statistically the same or whether they were different.

Field 2

Lease	Number of Measurements	Mean ($\overline{X}\phi$)	Variance ($S_{X\phi}^2$)
1	179	.1948	.000383
2	71	.1950	.000670
3	95	.2031	.000641
	345		

For leases:

$$Se^2 = .000512$$

$$S_p^{\ 2} = .002368$$

$$F = \frac{S_p^{\ 2}}{Se^2} = 4.62$$

Degrees of freedom with $Se^2 = 345 - 3 = 342$

Degrees of freedom with $S_p^{\ 2} = 3 - 1 = 2$

The F table gives:

$F = 3.03$ at the 5 percent confidence level

$F = 4.69$ at the 1 percent confidence level

Since the calculated F, $4.62 < 4.69$, the observed difference in the three leases means could reasonably be explained on the basis of the scatter of the observed data, and the three means are statistically the same.

It may be concluded from the two F tests that the individual well means are statistically different, but the lease means are statistically the same. The well means each

represent individual points in the field and since these points are different, heterogeneity exists between the various wells. As the wells are lumped together to form a lease mean the heterogeneity of the individual points are absorbed into larger somewhat homogeneous units, these larger units exhibiting similar statistical characteristics.

Modified Tukey Test

The results of the F test on the means of the 14 wells gave convincing evidence of differences among the means, but the F test gave no clue as to how many differences there were. In a group of a means there are in all a(a-1)/2 potential differences; 14(13)/2 = 91 among the wells. Does each mean differ from all the rest, or are some of them the same?

One method of investigating the differences is by the Tukey test (modified)[5]. The test is made by computing a difference, D, which is significant at the 5 percent level, then comparing it with the a(a-1)/2 sample differences. D is the product of $S_{\bar{X}}$ and a factor, Q, taken from a Q table[6] which is itself computed on the basis of the distribution of the deviations among compared means.

(5-11) $$S_{\bar{X}} = \frac{S_X}{\sqrt{n}}$$

and:

(5-12) $$D = Q\ S_{\bar{X}}$$

To determine Q enter the upper heading of the Q table with the number of treatments (14 wells) and degrees of freedom, f, for samples (331) indicated at the left of the table. From the table with 14 treatments and f = 331, Q is 4.74.

For porosity, Field 2,

$$S_{\overline{X}}^2 = 1.86 \times 10^{-5}$$

and,

$$S_{\overline{X}} = 4.313 \times 10^{-3}$$

Therefore:

$$D = (4.74)(4.313 \times 10^{-3}) = .02048$$

The differences to be compared with D are shown in Table IX.

In Table IX the $\overline{X}_{\phi}$ are arranged from high to low and each is subtracted from those above. Of the 91 differences, only 15, indicated by *, exceed D = .0205. One inference from the table is that the population represented by the mean of well 13 is different from that of wells 14, 8, 12, 4 and 11. Similar statements may be made about any particular well. For instance it may be inferred that the population represented by the mean of well 14 is different from that of wells 13, 7, 6, 2, 10, and 9 but the differences between well 14 and the remaining wells are not significant. This above procedure is therefore useful for seeking homogeneities and dissimilarities among wells. Such information could be valuable in explaining similarities and differences in individual well production behavior.

TABLE IX

Differences Between Well Means, Modified Tukey Test

Well	$\bar{X}$	$\bar{X}$-.181	$\bar{X}$-.184	$\bar{X}$-.186	$\bar{X}$-.189	$\bar{X}$-.190	$\bar{X}$-.191	$\bar{X}$-.199	$\bar{X}$-.200	$\bar{X}$-.203	$\bar{X}$-.205	$\bar{X}$-.208
14	.212	*.031	*.028	*.026	*.023	*.022	*.021	.013	.012	.009	.007	.004
8	.208	*.027	*.024	*.022	.019	.018	.017	.009	.008	.005	.003	
12	.208	*.027	*.024	*.022	.019	.018	.017	.009	.008	.005	.003	
4	.205	*.024	*.021	.019	.016	.015	.014	.006	.005	.002		
11	.203	*.022	.019	.017	.014	.013	.012	.004	.003			
3	.200	.019	.016	.014	.011	.010	.009	.001				
1	.199	.018	.015	.013	.010	.009	.008					
5	.199	.018	.015	.013	.010	.009	.008					
9	.191	.010	.007	.005	.002	.001						
10	.190	.009	.006	.004	.001							
2	.189	.008	.005	.003								
6	.186	.005	.002									
7	.184	.003										
13	.181											

Sequential Method of Testing

A sequential method of testing the differences between the means devised by Hartley[7] is a somewhat more powerful testing procedure than the modified Tukey method. For this test, not one Q is taken from the Q table but several, one for each range of the treatment means. For the well means, adjacent means in the array are tested with Q = 2.77 for a = 2; for two ranks apart in the array use Q = 3.32 for a = 3; for three ranks apart in the array use Q = 3.63 for a = 4; use Q = 4.74 only for the extreme range where a = 14. The corresponding D are:

a	Q	D
2	2.77	.012
3	3.32	.014
4	3.63	.016
5	3.86	.017
6	4.03	.0174
7	4.17	.018
8	4.29	.0185
9	4.39	.0189
10	4.47	.0193
11	4.55	.0196
12	4.62	.0201
13	4.68	.0202
14	4.74	.0205

These D are entered in the northeast-southwest diagonals of the table of differences, Table X; with the D's in parenthesis. Each difference is now compared with its own D, the difference being judged significant if it is larger than its D. A useful rule to be observed is that if any difference is less than its D then no further testing needs to be done to the right of that difference in its row or below it in the column.

TABLE X

Differences Between Well Means, Sequential Testing Method

Well	$\overline{X}$	$\overline{X}$-.181	$\overline{X}$-.184	$\overline{X}$-.186	$\overline{X}$-.189	$\overline{X}$-.190	$\overline{X}$-.191	$\overline{X}$-.1987	$\overline{X}$-.199	$\overline{X}$-.200	$\overline{X}$-.203
14	.212	*.031 (.0205)	*.028 (.0202)	*.026 (.0201)	*.023 (.0196)	*.022 (.0193)	*.021 (.0189)	.0133 (.0185)	.013	.012	.009
8	.208	*.027 (.0202)	*.024 (.0201)	*.022 (.0196)	.019 (.1093)	.018 (.0189)	.017 (.0185)	.0093	.009	.008	.005
12	.2077	*.0267 (.0201)	*.0237 (.0196)	*.0217 (.0193)	.0187 (.0189)	.0177 (.0185)	.0167	.009	.0087	.0077	.0047
4	.205	*.024 (.0196)	*.021 (.0193)	*.019 (.0189)	.016 (.0185)	.015	.014	.0063	.006	.005	.002
11	.203	*.022 (.0193)	*.019 (.0189)	*.017 (.0185)	.014	.013	.012	.0043	.004	.003	
3	.200	*.019 (.0189)	.016 (.0185)	.014	.011	.010	.009	.0013	.001		
1	.199	.018 (.0185)	.015 (.018)	.013	.010	.009	.008	.0003			
5	.1987	.0173 (.0180)	.0147 (.0174)	.0127	.0097	.0087	.0077				
9	.191	.010 (.0174)	.007 (.017)	.005	.002	.001					
10	.190	.009 (.017)	.006 (.016)	.004	.001						
2	.189	.008 (.016)	.005 (.014)	.003							
6	.186	.005 (.014)	.002 (.012)								
7	.184	.003 (.012)									
13	.181										

Well	$\overline{X}$-.205	$\overline{X}$-.2077	$\overline{X}$-.208
14	.007	.9943	.004
8	.003	.0003	
12	.0027		

From Table X, it is observed that there are 18 significant differences, marked by an *, which signifies that the sequential method detects a greater number of differences than the modified Tukey method.

From the table the wells similar to any given well may be detected. For example, it may be stated that the mean of well 9 is not significantly different from any of the other means except well 14. The mean porosity of well 3 is not significantly different from that of any other mean except for well 13.

The sequential testing method then provides a comparison mechanism for determining the relationship between any well mean desired with the other well means in the group.

Homogeneity of Variances

Several tests are available for testing the homogeneity of the variances of the several samples. The appropriate test depends upon the type of possible variation of the variances that is visualized. If the deviation from equality, i.e., $(\sigma_1^2 = \sigma_2^2 = \ldots. = \sigma_k^2)$ is conceived to be caused by a random variation then an appropriate variance homogeneity test is Bartlett's test[8] wherein the statistics B and C are computed where:

$$B = \frac{1}{C}(v \ln S^2 - \Sigma\, v_i \ln S_i^2)$$

(5-13)

$$= \frac{2.30259}{C}(v \log_{10} S^2 - \Sigma\, v_i \log_{10} S_i^2)$$

(5-14) $$C = 1 + \frac{\Sigma(\frac{1}{v_i}) - \frac{1}{v}}{3\ (K-1)}$$

where:

$$v = \Sigma\ v_i \text{ and } S^2 = \frac{\Sigma\ v_i\ S_i^{\ 2}}{v}$$

v_i is the degrees of freedom associated with each sample variance $S_i^{\ 2}$ and K is the number of variances being considered.

The statistic B is known to satisfactorily approximate the Chi-square (χ^2) distribution corresponding to K-1 degrees of freedom.[9] A calculated B value greater than the χ^2 value at the given degrees of freedom and confidence coefficient is evidence for rejecting the hypothesis $\sigma_1^{\ 2} = \sigma_2^{\ 2} = \ldots. = \sigma_k^{\ 2}$. Calculations for porosities of Field 2 are given below to illustrate the method.

TABLE XI

Computation of Bartlett's Test of Homogeneity of Variance-Porosity Field 2

Well No.	v_i (n-1)	$S_i^{\ 2}$ *	log $S_i^{\ 2}$	v_i log $S_i^{\ 2}$	$\frac{1}{v_i}$	$v_i\ S_i^{\ 2}$
1	23	1.983	.29732	6.8384	.044873	45.609
2	27	2.019	.30514	8.2388	.037037	54.513
3	25	1.283	.10823	2.7058	.040000	32.075
4	26	2.551	.39967	10.3914	.03846	66.326
5	17	2.633	.42537	7.2313	.05882	44.761
6	35	8.903	.94954	33.2339	.02857	311.605
7	13	1.292	.11126	1.4464	.07692	16.796
8	22	1.740	.24055	5.2921	.94546	38.280
9	33	6.071	.78326	25.8476	.03030	200.343
10	19	2.282	.35832	6.8081	.05263	43.353
11	30	4.023	.60455	18.1365	.03333	120.690
12	18	8.938	.95124	17.1223	.05555	160.884
13	14	9.458	.97580	13.6612	.07143	132.412
14	29	3.109	.49262	14.2860	.03448	90.161
	331			171.2396	.64788	1357.813

*Converting porosity data to percent (.199 to 19.9) changes variance from .0001983 to 1.983 percent.

$$K = 14 \quad v = 331$$

(5-13) $$B = \frac{2.30259}{C} (v \log S^2 - \Sigma v_i \log S_i^2)$$

(5-14) $$C = 1 + \frac{\Sigma(\frac{1}{v_i}) - \frac{1}{v}}{3\ (K-1)}; \quad \frac{1}{v} = \frac{1}{331} = .0030211$$

$$C = 1 + \frac{.64788 - .0030211}{3\ (14-1)}$$

$$C = 1 + .01433 = 1.01433$$

$$S^2 = \frac{\Sigma v_i S_i^2}{v} = \frac{1357.813}{331} = 4.1022$$

$$\log_{10} S^2 = .61302$$

(5-13) $$B = (\frac{2.30259}{1.01433})((331)(.61302) - 171.2396)$$

$$B = (2.27006)(31.67)$$

$$B = 71.98$$

From χ^2 table[10] with 13 degrees of freedom (14-1)

$\chi^2 = 22.4$ (at 5 percent level)

$\chi^2 = 27.7$ (at 1 percent level)

$\chi^2 = 29.8$ (at 1/2 percent level)

A χ^2 of 29.8 at the 1/2 percent level indicates that there is only 1 chance out of 200 that the differences between the variances could be caused by random variation if the calculated $B > 29.8$. Since the $B = 71.98$, it may be concluded that the differences between the variances of the 14 wells are significant and cannot be accounted for by

random variations. Therefore the hypothesis that $\sigma_1^2 = \sigma_2^2 = \ldots. \sigma_k^2$ is rejected.

Summary

Various statistical tests that may be used to analyze core samples by use of the core means and variances have been illustrated. A core sample provides certain information about the population from which it is extracted. The respective populations sampled by $core_1$, $core_2$, ... $core_{14}$, for Field 2 may be thought of as falling into one of four categories:

1. The 14 samples represent identical populations, the apparent differences resulting from sampling error.
2. They represent identical variances but have different means.
3. They exhibit identical means but different variances.
4. They are totally different.

The F test indicated that the populations had different means. Bartlett's test showed that the populations had different variances. Therefore we may conclude that the populations represented by the 14 samples are different.

By use of the Tukey test, and the Sequential test it is possible to determine what sub-groupings of the 14 populations are statistically homogeneous. An indication that any particular well is statistically the same as a group of other wells should prove to be a valuable tool in correlating the expected performance of that well with the

performance of the group with which it may be identified.

In this chapter particular attention has been given to the distribution of sample means and of sample variances, yet in reality there is very little exact knowledge, based upon sedimentation theory, as to the appropriate distributions to expect. It is to be hoped that the methods of the present chapter will be useful for quantitatively characterizing these distributions and thereby aid in the formulation of more precise theories of sedimentation.

CHAPTER VI
SAMPLING

Introduction

The prime purpose of core analysis is to gain some understanding of the characteristics of the field properties under consideration, namely porosity, permeability, and fluid saturations. As previously discussed, once sufficient information becomes available, it may be used to determine descriptive parameters for a field. The parameters may be expressed simply in terms of averages and deviations from the averages. Or more extensive parameters may be developed in terms of theoretical frequency curves and the moment parameters α_3^2 and δ.

Considering the initial period in the life of a field before an extensive amount of data becomes available, the question arises as to what inferences can be derived from the data as it becomes available.

Sampling

Sampling may be defined as the selection of part of an aggregate or population, on the basis of which a judgement or inference about the aggregate or population is made.[1] This sampling theory is a study of relationships existing between a population and samples drawn from the population. Statisticians have long worked within the problem of reconstruction of a universe of variables by means of samples that comprise a small percentage of the universal or population from which the samples were drawn. Core analysis data appears

to comply with the requisite of random sampling as stipulated by theoretical statistics.

Sampling theory is useful in estimation of the unknown population quantities such as population mean, standard deviation, variance, etc., referred to as population parameters, from a knowledge of corresponding sample quantities.

In general a study of inferences made concerning a population by use of samples drawn from it together with indications of the accuracy of such inferences using probability theory, is called statistical inference.

Considering a core analysis as a random sample drawn from an infinite population, the sub-surface of the earth, what inferences can be made about the population? From the core analysis there may be obtained a sample mean and a sample standard deviation where:

(6-1)
$$\overline{X} = \frac{\Sigma X}{n}$$

(6-2)
$$S_X = \frac{\sqrt{\Sigma(X-\overline{X})^2}}{n}$$

The sample mean, which serves as an estimate of the population mean will be of chief interest in considering a core sample. An important theorem in statistics states "that for almost all populations the probability distribution of the sample mean based upon a simple random sample will be an approximately normal one if the sample size is sufficiently large."[2] The standard deviation of the probability distribution of the sample mean, denoted by $\sigma_{\overline{X}}$ is calculated by:[3]

(6-3)
$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

where σ_X is the standard deviation of the population being sampled, and n is the sample size.

Therefore if the standard deviation of the population is known, $\sigma_{\overline{X}}$ may be calculated. However, after only one core analysis, the standard deviation of the population is not known and an estimate of this standard deviation must be made. A point estimate of this population characteristic is made from the sample standard deviation, S_X, and the sample mean is the best point estimate available for the population mean.

To illustrate the application of point estimates, the numerical data obtained from an assumed first core analysis taken from Field 2, Figure 28, will be used.

From the first core analysis, well 1, we obtain:

$$\overline{X}_\phi = 18.6\%$$
$$S_{X\phi} = 2.98$$
$$\overline{X}_{so} = 45.8\%$$
$$S_{Xso} = 4.56$$
$$\overline{X}_{sw} = 45.3\%$$
$$S_{Xsw} = 4.14$$
$$n = 36$$

Therefore the best estimate of the various population means and standard deviations are:

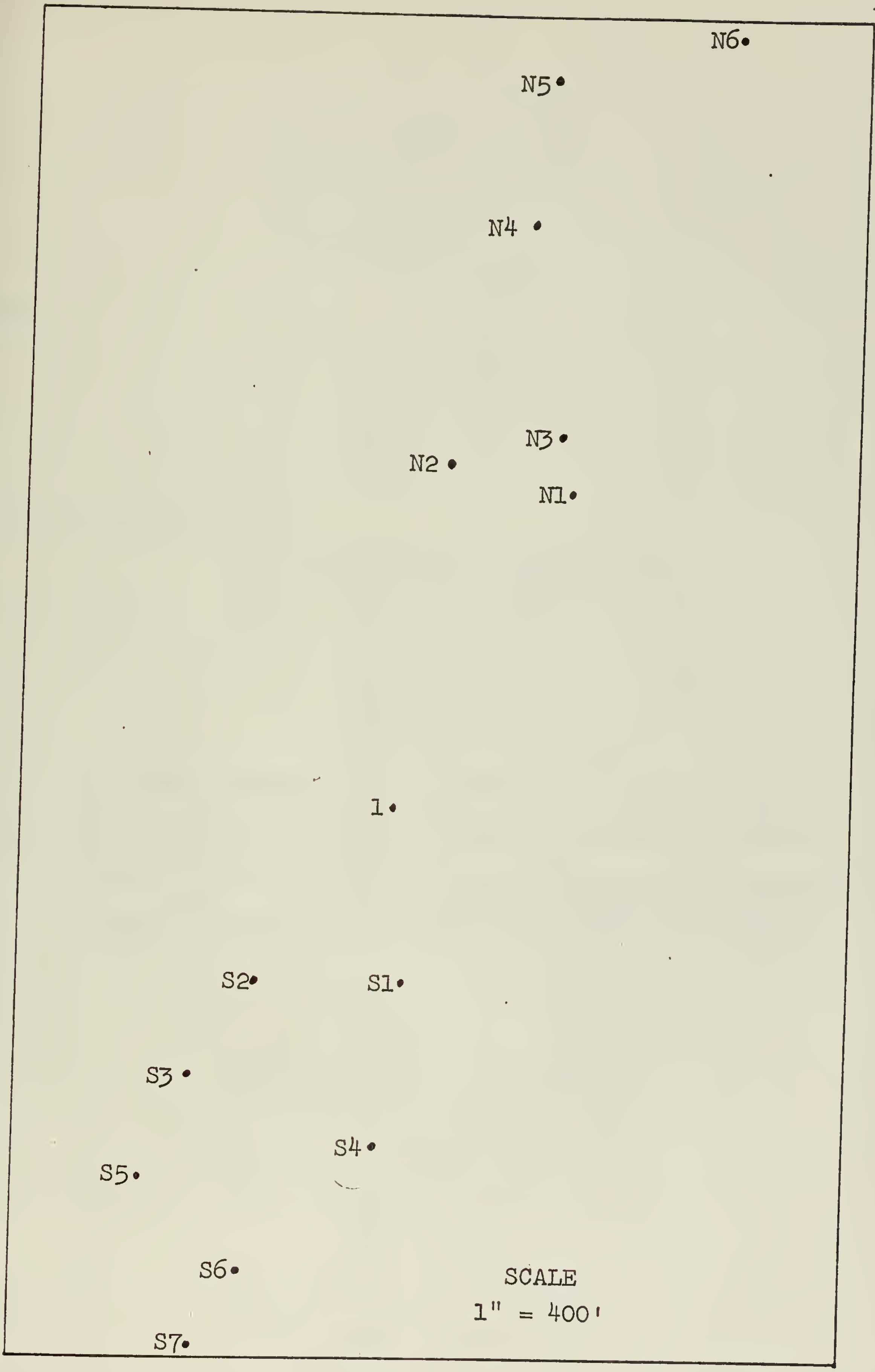


FIGURE 28 MAP OF WELLS CORED--FIELD 2

$$\overline{X}_{\phi\text{population}} = 18.6\%$$

$$\overline{X}_{so_{\text{population}}} = 45.8\%$$

$$\overline{X}_{sw_{\text{population}}} = 45.3\%$$

and,

$$\sigma_{X\phi} = 2.98$$

$$\sigma_{Xso} = 4.56$$

$$\sigma_{Xsw} = 4.14$$

And, we assume that:

$$\sigma_{\overline{X}\phi} = \frac{\sigma_{X \text{ estimate}}}{\sqrt{n}} = \frac{2.98}{\sqrt{36}} = .48$$

$$\sigma_{\overline{X}so} = .76$$

$$\sigma_{\overline{X}sw} = .69$$

For normal probability distributions, the area under the normal distribution curve between the mean $\pm$ 2 standard deviations is about .95 out of a total area of 1.[4] Therefore the following limits may be constructed with a confidence of .95.

(6-4) $$\overline{X} \pm 2\,\sigma\,\overline{X}$$

$$\overline{X}_{\phi} \pm 2(.48) = 18.6 \pm .96$$

$$17.64 - 19.56$$

$$\overline{X}_{so} \pm 2(.76) = 45.8 \pm 1.52$$

$$44.2 - 47.4$$

$$\overline{X}_{sw} \pm 2(.69) = 45.3 \pm 1.38$$

$$43.9 - 46.7$$

We can conclude with a .95 probability that:

$\bar{X}_{\phi}$ for Field 2 is somewhere between 17.64 and 19.56%.

$\bar{X}_{so}$ for Field 2 is somewhere between 44.2 and 47.4%.

$\bar{X}_{sw}$ for Field 2 is somewhere between 43.9 and 46.7%.

The above conclusions are based on knowledge of the true value of σ_X which was not the case. Statistical history has indicated though that if the sample size is reasonably large, $n > 30$, the S_X may be used as an estimate of σ_X without materially changing the reliability of the above conclusions.

Control Charts

From the previous section, estimates of the probable range of the various property means of Field 2 were determined. The ranges were based upon one core analysis which was taken from one point in the field. For the population which this core sample represents, the sample mean ranges determined should prove to be satisfactory range estimates of the population mean. As drilling moves away from this point, a different areal population may be encountered with different characteristics from those determined by the initial core sample. In this case it would be expected that the new population characteristics may deviate from those estimated from the initial core sample.

In statistical quality control procedures for production processes, a control chart is used to indicate when a process has changed or "gone out of control" and is no longer

producing within prescribed specifications. Basically the production control procedures consist of computing a process mean and standard deviation, constructing a control chart within the limits of $\overline{X}$process $\pm$ 3 $\sigma_{\overline{X}}$, taking periodic samples from the production line and computing the mean of the samples taken. If the mean falls between the limits set on the control chart, it is concluded that the process has not changed or "is in control." If the mean does not fall within the limits, it is concluded that the process has changed or "is out of control."[5]

Core sampling may be considered analogous to production line sampling with areal sampling of populations considered to be the analog of periodic process sampling. As core samples are taken at varying distances from the initial sample, the means may be plotted on a control chart to indicate if the population characteristics are still "under control," or if not, whether a new areal population has been encountered. For a control chart, $\pm$ 3 $\sigma_{\overline{X}}$ is normally used. Thus: (from previous section)

$$\overline{X}_{\phi} \pm 3\ \sigma_{\overline{X}} = 18.6 \pm 3(.48)$$
$$18.6 \pm 1.44$$
$$17.16 - 20.04$$

$$\overline{X}_{so} \pm 3\ \sigma_{\overline{X}} = 45.8 \pm 3(.76)$$
$$45.8 \pm 2.28$$
$$43.52 - 48.08$$

$$\overline{X}_{sw} \pm 3\ \sigma_{\overline{X}} = 45.3 \pm 3(.69)$$

$$45.3 \pm 2.07$$

$$43.23 - 47.37$$

Using the above, a control chart may be constructed. From Figure 28 it can be observed that the wells are located in a general north-south direction with well 1 located approximately at the center of the field.

For this particular geographical arrangement, two control charts to test the hypothesis that the mean of the field falls within the probable range are constructed, one for moving in a northerly direction, one for the southerly direction.

For moving North:

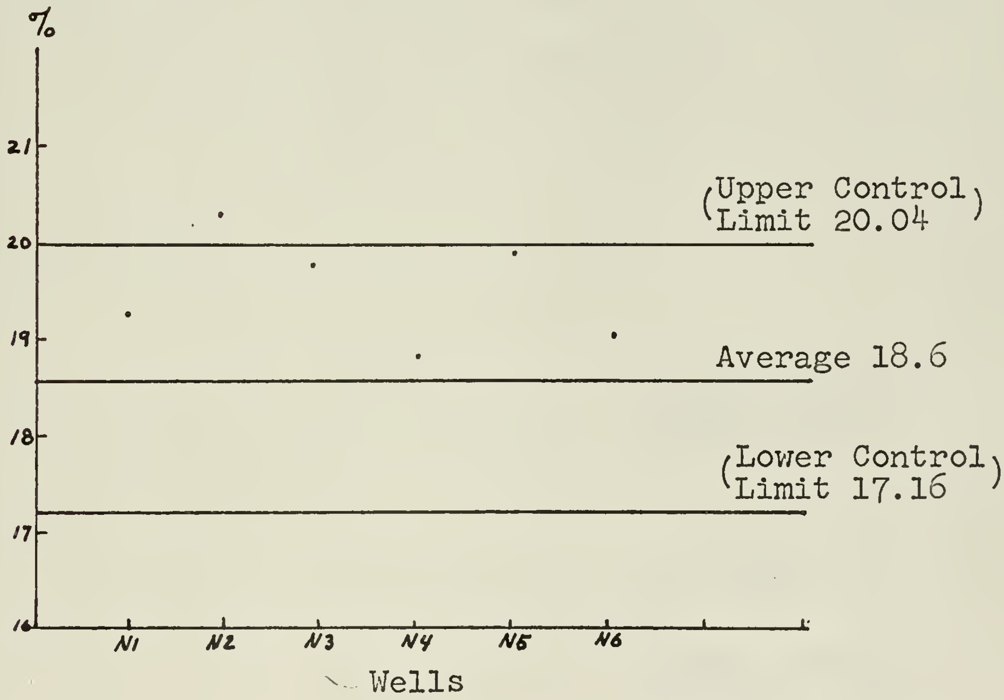


FIGURE 29 Control Chart for Average Porosity

The core analysis from well N1 has a $\overline{X}_\phi$ of 19.2. This is plotted on the control chart, and falls within the limits set. It is concluded that the sampling is still under control and the population has not changed from that indicated by the initial sample. As the core analysis from well N2 through N6 are taken, their means are plotted on Figure 29.

Moving north from initial well:

Well	$\overline{X}_\phi$	
N1	19.2	under control
N2	20.5	possibly out of control
N3	19.9	under control
N4	18.9	under control
N5	19.9	under control
N6	19.0	under control

It appears that the probable range of the field mean calculated from the initial core analysis gives a good representation of the field mean from the initial point to well N6.

For moving South:

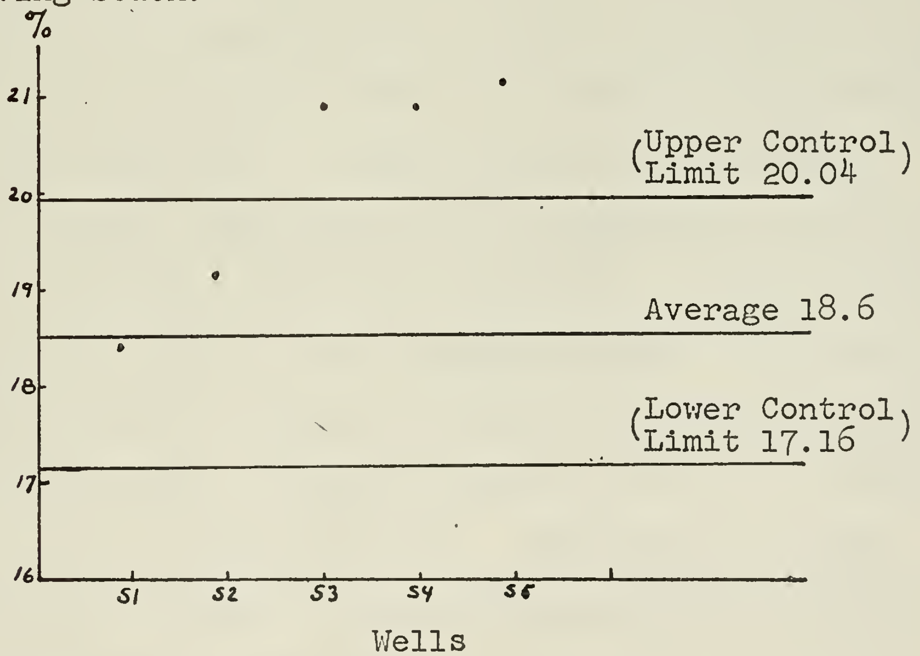


FIGURE 30 Control Chart for Average Porosity

Moving south from initial well:

Well	$\overline{X}_{\phi}$	
S1	18.4	under control
S2	19.1	under control
S3	20.8	possibly out of control
S4	20.8	more indication of out of control
S5	21.2	out of control

For movement in a southerly direction the initial conclusion concerning the field mean proved correct from the initial point to well S2. When it becomes apparent that the initial hypothesis concerning the mean is not tenable, in this case after well S4 was drilled, a new probable range of the areal mean should be calculated from the data obtained from well S3 and S4. This new probable range would then be the best estimate of the population mean for the areal population from S4 southward.

The adoption of the above method, which is based on production quality control techniques, to the analysis of reservoirs is one possible way in which a quantitative criterion could be used to segment a field into smaller homogeneous units. Many extensions to this procedure are possible. These involve the comparison of the segmentation created by use of the different reservoir properties. In this regard it is possible that the areal segments would not coincide when using different properties such as porosity, permeability, or fluid saturations. However there is also no "a priori" reason for their being different. Similarity of areas would probably indicate a fairly high degree of correlation between the variables involved.

<u>Confidence Interval for Population Proportion</u>

If a random sample size is large, $n > 30$ usually being considered sufficiently large, a confidence interval for the population proportion can be constructed in a similar way as for the population mean, since the probability distribution of the sample proportion will be assumed to be approximately normal.

The standard deviation of the probability distribution of the sample proportion can be estimated from:[6]

$$S_{\bar{p}} = \frac{\sqrt{\bar{p}(1-\bar{p})}}{n-1} \tag{6-5}$$

where $\bar{p}$ is the sample proportion and $S_{\bar{p}}$ denotes the estimate of the true standard deviation $\sigma_{\bar{p}}$.

The confidence interval for the population proportion is:

$$\bar{p} \pm Z\, S_{\bar{p}} \tag{6-6}$$

where Z is the normal deviate corresponding to the desired confidence coefficient.[7]

To illustrate the application of confidence intervals for population proportions to core analysis, suppose that it has been decided that an oil saturation of 30 percent or less is undesirable in considering the amount of pay zone in a field. The initial core analysis taken from Field 2 will be used for example calculations.

There were 36 samples from well 1, 6 of which were ≤ 30 percent. Thus:

$$\bar{p} = \frac{6}{36} = \frac{1}{6} = .167$$

and,

$$S_{\bar{p}} = \frac{\sqrt{(.167)(.833)}}{35} = .066$$

And for a confidence coefficient of .95, Z is 1.96,[8]

$$\bar{p} \pm Z\ S_{\bar{p}}$$

$$.167 \pm 1.96\ (.066)$$

$$.167 \pm .129$$

Thus it may be concluded with a confidence coefficient of 95 percent that the percentage of the field pay zone that is undesirable according to the predetermined criteria is somewhere between 3.8 percent and 29.6 percent.

Another way to interpret the use of the population proportion and its relation to the confidence coefficient is illustrated by the following modification of the above problem. Instead of establishing a confidence coefficient of 0.95 and determining the range in the percent of formation which would be unproductive, suppose that one seeks to know the confidence coefficient associated with a particular percentage range such as 1/6 ± 3 percent, or .167 ± .03. One then determines the value of Z corresponding to the ± 3 percent range. In this case:

$$\text{interval in population proportion} = \bar{p} \pm Z\ S_{\bar{p}}$$

A value of Z is sought such that $Z(S_{\bar{p}}) = .03$ or $Z = .03/.066 = .455$. This value of Z corresponds to a confidence

coefficient of .35. The interpretation is that the true unproductive fraction of the formation has 2 chances out of 3 to be outside of the proportion .167 $\pm$ .03 or a 1 chance in three (approximately 35 percent) of lying within the proportion range .167 $\pm$ .03.

Summary

Two rather elementary techniques of estimating information about a field when the amount of core analysis data is limited, in this case one well, have been presented. For Field 2, the sample mean of well 1 gave good results as an estimation of the population mean for the areal population from well S2 to N6.

The utility of these sampling techniques can only be confirmed after being tried and tested with a large number of fields where the actual results could be compared with the estimated predictions. It would be conceivable to assume that over the long run, sampling techniques would provide a better estimate of what is to be expected than those estimates determined intuitively. Granted, in some cases the reservoir predictions by sampling techniques would prove to be in error, and yet the estimates could show less error than the error or difference found by intuitive predictions.

Sampling results should be recognized as only estimates and not as infalliable predictions. Yet the sampling estimates should prove better than no estimates at all. For any one reservoir the estimates are either right or

wrong. The .95 confidence coefficient means that if the procedure is followed for a long enough time over a sufficiently large number of reservoirs, the results should be correct 95 percent of the time.

It was the purpose of this discussion on sampling to indicate how certain techniques could be applied and possibly to stimulate the study of the applicability of other more advanced or specific sampling techniques. It is felt that the work and investigation in this area has just begun with unlimited possibilities yet to be explored.

CHAPTER VII
SUMMARY AND CONCLUSIONS

Introduction

At the outset of this study two alternate courses of investigation were considered for the statistical analysis of the field data available; the first, to concentrate on one statistical technique and analyze it in detail, and second, to explore various techniques with less detailed analysis of any one point. The latter course was chosen with three very broad areas included in the study, i.e., generalized or skewed distribution curves including transformation of data to logrithms for analysis, analysis of variances, and sampling procedures.

The choice of this latter course was prompted by the realization that relatively little use of theoretical statistics has been made in core analysis as judged by available discussion in the petroleum literature. An attempt was made to present in one volume the necessary calculation procedures and interpretations to serve as a broad guide for a statistical analysis of field data. Additionally, it is hoped that the study will serve as a basis for further investigation into the applicability of the science of statistics to reservoir problems.

Summary

This thesis has attempted to indicate the applicability of certain statistical procedures to the analysis of core samples. Statistical techniques were employed to segregate

relevant and meaningful information from a mass of raw data.

A method of determining and describing frequency distributions of the physical properties of a reservoir is presented in Chapter III. The Pearson family of generalized frequency curves provides an effective means of classifying observed frequency distributions without being restricted to a few classical distributions such as the normal curve, the Poisson curve or the binomial distribution.

Chapter IV is devoted to the study of the effects of the transformation of raw data to logrithms for analysis. The results obtained were compared with the results of Jan Law's[1] work.

Analysis of variances is studied in Chapter V with various testing procedures shown. The F test and Bartlett's test are employed to test for uniformity of the population represented by the various well samples. The modified Tukey and Sequential tests are used to determine what subgrouping of the wells are statistically homogeneous.

Sampling procedures are presented in Chapter VI. Sample means are analyzed and presented on control charts by techniques analogous to production line sampling techniques. Population proportions are presented with confidence intervals.

The following listing presents a summary of the main equations associated with the above statistical procedures.

For Frequency Distributions

Pearson Type I Curve

(2-24) $$Y = C(t-r_1)^{m_1} (r_2-t)^{m_2}$$

Type IV Curve

(2-31) $$Y = C[(t+r)^2 + S^2]^{-m} e^{-v \tan^{-1}\left(\frac{t+r}{S}\right)} e^{v^{\frac{\pi}{2}}}$$

Type VI Curve

(2-34) $$Y = C(Z^{m_2})(Z-\alpha)^{m_1}$$

Normal Curve

(2-40) $$Y = C\ e^{-\frac{t^2}{2}}$$

Type III Curve

(2-43) $$Y = C_1(A+t)^{A^2-1}\ e^{-At}$$

For Analysis of Variance

The F Test

(5-10) $$F = \frac{S_p^{\ 2}}{Se^2}$$

where:

(5-9) $$S_p^{\ 2} = n_o\ S^2_{m,p}$$

and,

(5-3) $$Se^2 = \frac{\Sigma^{n_1}(X_1-\overline{X}_1)^2 + \ldots.\ \Sigma^{n_k}(X_k-\overline{X}_k)^2}{(n_1 + \ldots.\ n_k) - n_i}$$

Modified Tukey and Sequential Testing

(5-11) $$S_{\overline{X}} = \frac{S_X}{\sqrt{n}}$$

(5-12) $$D = Q\ S_{\overline{X}}$$

Bartlett's Test

(5-13) $$B = \frac{1}{C}(v \ln S^2 - \Sigma v_i \ln S_i{}^2)$$

where:

(5-14) $$C = 1 + \frac{\Sigma \frac{1}{v_i} - \frac{1}{v}}{3\ (K-1)}$$

For Sampling

(6-2) $$S_X = \frac{\sqrt{\Sigma(X-\overline{X})^2}}{n}$$

(6-3) $$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

(6-5) $$S_{\overline{p}} = \frac{\sqrt{\overline{p}(1-\overline{p})}}{n-1}$$

(6-6) $$\overline{p} = \pm Z\ S_{\overline{p}}$$

Conclusions

In conjunction with the comments given in the summary sections of the specific chapters, the major conclusions to be derived from the data studied are believed to be:

1. Generalized frequency curves are useful in reflecting the variation of individual properties throughout a single depositional unit, and in characterizing and distinguishing the unit from other units in the same geological basin. The terms α_3 and δ should provide additional relevant numerical parameters for mathematical model analysis of a reservoir.
2. Logrithmic transformation is occasionally useful to remove some skewness from permeability data, but does not necessarily convert the data to the normal distribution under all circumstances.
3. Analysis of variances tests provide a useful tool to investigate homogeneity between well properties within a reservoir or within sub areas of a reservoir.
4. Sampling techniques may be utilized to estimate unknown population quantities from a knowledge of relatively small sample quantities.
5. The science of theoretical statistics applied to reservoir analysis is in its infancy and appears to offer many opportunities for the solution of reservoir problems.

Recommendations for Further Investigation

While working with the Pearson frequency curves, certain areas of further investigation appeared to offer promise:

1. An investigation into the relationship between reservoir performance and the type of frequency distribution followed by its physical properties. For example, does a field with a Pearson Type III S_o distribution perform differently than a field with a Type I, other factors being the same? What is the correlation, if any, between the parameters (α_3, δ) and field performance, the parameters α_3 and δ being considered for any physical property or group of properties?
2. Sampling techniques are available for populations that approximate the normal frequency distribution. Could reliable techniques be developed for populations that satisfy certain types of the Pearson family of frequency distributions? A table of Areas Under the Type III_B Curve has been developed[2] and could possibly serve as a point of departure for investigating sample techniques for core samples.

Numerous tests have been devised for the analysis of variances. Certain of these tests were presented in this thesis. Other tests may prove more useful than the ones presented. Further investigation into this area of statistical analysis is highly recommended.

NOTES

Chapter I

1. Mode, E. B., Elements of Statistics, Englewood Cliffs: 1951, p. 1.

2. Shewhart, W. A., Economic Control of Quality of Manufactured Product, New York: 1931, pp. 136-137.

   Shewhart is original source of the Table, however this Table is a copy of Table 1 from,

   Wedel, A. M., "Karl Pearson's System of Generalized Frequency Curves", Unpublished Master's Thesis, University of Kansas, 1948, p. 5.

3. "Coring and Sampling", The Oil and Gas Journal, 57:C29, January 28, 1959.

4. The data summarized in Table II was obtained from 181 laboratory reports of the analysis of core samples prepared by the Oil Field Research Laboratories, Chanute, Kansas.

Chapter II

1. Carver, H. C., Handbook of Mathematical Statistics, Cambridge, 1924, p. 103.

2. Wedel, op cit, p. 31.

3. Carver, op. cit., p. 103.

4. Ibid, p. 104.

5. Craig, C. C., "A New Exposition and Chart for the Pearson System of Frequency Curves", Annals of Mathematical Statistics, 1936, Vol. 7, p. 17.

6. Carver, op. cit., p. 104.

7. Ibid.

8. Hart, W. L., College Algebra, New York, p. 338.

   "If f(x) and g(x) are polynomials, then f(x)/g(x) is called a rational fraction. A rational fraction is called a proper fraction if the degree of the numerator is less than the degree of the denominator. In advanced mathematics it is sometimes necessary to decompose a given proper fraction into a sum of more simple fractions, called partial fractions." Hart shows how the proper fraction,

$1/b_2 \left[\frac{a - t}{(t-r_1)(t-r_2)}\right]$ may be reduced to

$$1/b_2 \left[\frac{A}{t-r_1} + \frac{B}{t-r_2}\right]$$

9. Craig, op. cit., p. 19.

10. Wedel, op. cit., p. 136.

11. Whittaker, E. and Watson, G., Modern Analysis, New York: 1946, p. 253, para. 12.4.

12. Ibid., p. 254, para. 12.41.

From The Expression of the Eulerian Integral of the First Kind in terms of the Gamma Function, the theorem is established that:

$$\beta(m,n) = \frac{\Gamma m \ \Gamma n}{\Gamma(m+n)}$$

13. The total probability was set equal to unity to determine C. For calculational purposes, the value N is placed in the numerator of the equation in order to account for all frequencies in a set of data. N = total number of measurements in a set of data.

14. Craig, op. cit., p. 20.

15. Ibid, p. 21.

16. Ibid, p. 22.

17. Craig, op. cit., pp. 23-24.

Craig in developing the Pearson equations in terms of α_3^2 and δ, shows this equation but does not offer a proof. For purposes of this study, the equations for the various constants, C, were used as presented by Craig.

18. Pearson, Karl, Tables for Statisticians and Biometricians, Cambridge: 1914, pp. 126-142.

19. Scarborough, J. B., Numerical Mathematical Analysis, Baltimore: 1950, p. 145.

and

Lowan, A. N., Bulletin of the American Mathematical Society, Oct., 1942, Vol. 48, No. 10, pp. 739-743.

Scarborough shows the method of numerical integration involving Gaussian coefficients which is considered to be the best general purpose quadrature procedure. The method for determining I where:

$$I = \int_a^b f(x)$$

by this technique is:

$$I = (b-a) \sum_{i=1}^{i=n} R_i f(Z_i)$$

where R_i are the selected Gaussian coefficients and the Z_i are selected, known values of the independent variable x transformed to a scale wherein the range a to b becomes the range from 0 to 1. For the present integration it was found necessary to use n = 16 for appropriate accuracy of the function. The "Weight Coefficients" for the Gauss Quadrature Formula for order 16 were taken from Lowan, pp. 739-743.

20. Smith, J. G., Elementary Statistics, New York, 1934, p. 287.

21. Carver, op. cit., p. 105.

22. Craig, op. cit., p. 27.

Chapter III

1. When data are grouped in discrete intervals, the moments are obtained by a summation process. The frequencies are considered to be lumped at the mid intervals which may not be the actual case. Therefore grouping errors may be introduced. An extensive discussion of a method of correcting for these grouping errors, known as Sheppard's corrections, is presented in Kendall, M.G, The Advanced Theory of Statistics, London, Vol. I, 1945, p. 71. Sheppard's corrections are:

$$X_2 = \mu_2 - 1/12,\ X_3 = \mu_3,\ X_4 = \mu_4 - 1/2\ \mu_2 + 7/240$$

2. Kendall, M. G., The Advanced Theory of Statistics, London, 1945, Vol. I, p. 49.

3. Ibid, p. 50.

4. Davis, H. T., Tables of Higher Mathematical Functions, Vol. I, Bloomington, 1933, p. 179.

5. *Ibid.*

6. *Ibid*, pp. 217-239.

7. *Ibid*. p. 181ff, gives this formula and other asymtatic expansions of the Gamma Function.

8. Scarborough, *op. cit.*, p. 132.

9. Arkin, H. and Colton, R., *Tables for Statisticians*, New York, 1956, p. 121.

also

Spiegel, M. R., Schoums Outline Series, *Theory and Problems of Statistics*, New York: 1961, p. 345.

10. Hoel, P. G., *Introduction to Mathematical Statistics*, New York, 1947, p. 188.

11. Law, Jan, *A Statistical Approach to the Interstitial Heterogeneity of Sand Reservoirs*, AIME Transactions, Vol. 155, 1944, p. 221.

Chapter IV

1. Law, Jan, *op. cit.*, p. 221.

2. Croxton, F. E. and Cowden, D. J., *Applied General Statistics*, Englewood Cliffs, 1959, p. 615.

Chapter V

1. Snedecor, George W., *Statistical Methods*, p. 244, Iowa State College Press, Ames, Iowa, 1957.

2. Mickley, Harold S., *Applied Mathematics In Chemical Engineering*, p. 73, McGraw-Hill, New York, 1957.

3. Mickley, *op. cit.*, p. 75.

4. Mickley, *op. cit.*, p. 76.
Snedecor, *op. cit.*, p. 245.

5. Snedecor, op. cit., p. 251.

6. *Ibid*, p. 252.

7. *Ibid*, p. 253.

8. Bennett, C., and Franklin, N., *Statistical Analysis in Chemistry and the Chemical Industry*, New York, 1954, p. 196.

9. *Ibid*, 197.

10. Arkin and Colton, *op. cit.*, p. 121.
also
Spiegel, *op. cit.*, p. 345.

Chapter VI

1. Neter, J., and Wisserman, W., *Fundamental Statistics for Business and Economics*, Boston: 1957., p. 267.

2. *Ibid*, p. 285.

3. *Ibid*, p. 310.

4. *Ibid*, p. 239.

5. *Ibid*, p. 376.

6. *Ibid*, p. 342.

7. *Ibid*, p. 343.

8. *Ibid*, also Spiegel, *op. cit.*, p. 343.

Chapter VII

1. Law, Jan, *op. cit.*, pp. 202-221.

2. Salvosa, L. R., "Table of Pearson's Type III Function", *Annals of Mathematical Statistics*, Ann Arbor: May 1930, p. 191.

BIBLIOGRAPHY

Arkin, H. and Colton, R., Tables for Statisticians, New York: Barnes and Noble, Inc., 1950.

Bennett, C., and Franklin, N., Statistical Analysis in Chemistry and the Chemical Industry, New York: John Wiley and Sons, 1954.

Carver, H. C., "Frequency Curves", Handbook of Mathematical Statistics, Cambridge: Riverside, 1924, 92-119.

"Coring and Sampling", The Oil and Gas Journal 57: C29, January 28, 1959.

Craig, C. C., "A New Exposition and Chart for the Pearson System of Frequency Curves", Annals of Mathematical Statistics, Vol. 7 (1936), 16-29.

Croxton, F. E. and Cowden, D. J., Applied General Statistics, Englewood Cliffs: Prentice-Hall, Inc., 1959.

Davis, Harold T., Tables of Higher Mathematical Functions, Bloomington, Ind.: Principia Press, Inc., 1933, Vol. 1.

Elderton, W. P., Frequency Curves and Correlation, Cambridge: University Press, 1938.

Hart, William L., College Algebra, New York: D. C. Heath, 1924.

Hoel, P. G., Introduction to Mathematical Statistics, New York: John Wiley and Sons, 1947.

Kendall, M. G., The Advanced Theory of Statistics, London: Charles Griffin and Co., 1945, Vol. 1.

Kendall, M. G., The Advanced Theory of Statistics, London: Charles Griffin and Co., 1946, Vol. 11.

Kendall, M. G., The Advanced Theory of Statistics, London: Charles Griffin and Co., 1951, Vol. 11.

Law, Jan, "A Statistical Approach to the Interstitial Heterogeneity of Sand Reservoirs", AIME Transactions, Vol. 155 (1944) pp. 202-221.

Lowan, A. N., Bulletin of the American Mathematical Society, Oct., 1942, Vol. 48, No. 10.

Mickley, Harold S., Applied Mathematics in Chemical Engineering, New York: McGraw-Hill, 1957.

Mode, E. B., Elements of Statistics, Englewood Cliffs: Prentice-Hall, Inc., 1951.

Neter, John and Wasserman, William, Fundamental Statistics for Business and Economics, Boston: Allyn and Bacon, Inc., 1956.

Pearson, Karl, Tables for Statisticians and Biometricians, London: Cambridge University Press, 1914.

Salvosa, L. R., "Tables of Pearson's Type III Function", Annals of Mathematical Statistics, Ann Arbor: Edwards Brothers, May 1930, p. [illegible].

Scarborough, J. B., Numerical Mathematical Analysis, Baltimore: John Hopkins Press, 1950.

Shewhart, W. A., Economic Control of Quality of Manufactured Product, New York: Van Nostrand, 1931.

Smith, James G., Elementary Statistics, New York: Henry Holt and Co., 1934.

Snedecor, George W., Statistical Methods, 5th Ed., Ames, Iowa: The Iowa State College Press, 1956.

Spiegel, Murray R., Schaum's Outline Series, Theory and Problems of Statistics, New York: Schaum Publishing Co., 1961.

Wedel, Arnold M., "Karl Pearson's System of Generalized Frequency Curves", Unpublished Master's thesis, The University of Kansas, Lawrence, Kansas, 1948.

Whittaker, E. T., and Watson, G. N., Modern Analysis, New York: The MacMillan Co., 1946.

APPENDIX A

MATHEMATICAL, FORTRAN, AND TECHNICAL NOTATIONS AND TERMS

Mathematical Notation	Fortran Notation	Description
A	A	A parameter of the Pearson Type III Curve, equal to 2/ALP3
A^2	ASQ	Square of A
a		Parameter of the differential equation (2-1); also skewness of the distribution of a set of data defined by equation (2-13); also, a term used in the Sequential method of testing to describe the range of means in an array
	AJ	Number of measurements in a particular class interval
	AK	Total number of measurements
α_n		The n^{th} α term where $\alpha_n = \frac{n^{th} \text{ moment above mean}}{\sigma^n}$, see equation (2-7)
α	ALPHA	A parameter of the Type VI Pearson Curve, see equation (2-37)
α_3	ALP3	A basic parameter of the Pearson System of Frequency Curves, see equation (2-7A)
${\alpha_3}^2$	ALP3S	Square of ALP3
α_4	ALP4	A basic parameter of the Pearson System of Frequency Curves, see equation (2-7A)
	AN	Number of class intervals
	ARG	Intermediate variable for computing YBAR of the type curve
	ARG1	Intermediate variable for computing YBAR of the type curve

Mathematical Notation	Fortran Notation	Description
B		A statistic calculated in Bartlett's test for variance homogeneity, see equation (5-13)
β		An intermediate value for derivation of the constant for Type I curve, see equation (2-28)
b		b_o, b_1, b_2 are terms of equation (2-2), where f(t) is expanded in a converging power series, values of terms are defined in equation (2-11)
	BELSQ	Summed BSQ between observed data and Normal Curve
	BSQ	Equivalent to an individual $\frac{(f-F)^2}{F}$ on page 45 for Normal Curve
C		A constant in the frequency curve for a particular set of data, also a statistic calculated in Bartlett's test, see equation (5-14)
	CHI	Equivalent to an individual $\frac{(f-F)^2}{F}$ on page 45 for Pearson Curve
χ^2	CHISQ	Summed CHI between observed data and Pearson Curve--a statistical distribution used for certain statistical testing
	CON	Storage locations for all parameters and the constant characteristic of a given Pearson Type Curve

Pearson Type	I, VI	III	IV
CON(1)	EM1	A	R
(2)	EM2	C	EM
(3)	R1		S
(4)	R2		V
(5)	C		C

	CST	The Gaussian coefficients (16) for integration in determining F(R,V), see Ref. 19 Chapter 3 for source of coefficients
D	D	An intermediate parameter for all main Pearson Type Curves, see equation (2-15)

Mathematical Notation	Fortran Notation	Description
D		For analysis of variance, a computed difference between means that is significant at a chosen level, see equation (5-12)
$\overline{X}$	DAVG	Arithmetic mean or average of data, see equation (6-11)
	DD(I)	Midpoint value for class interval I
	DECM	Difference between the highest value permissable in a given interval and the lowest permissable value in the next higher interval
	DELD	Difference between maximum and minimum value of a class interval
δ	DELTA	A basic parameter of the Pearson System of Frequency Curves, see equation (2-12)
	DMAX	Maximum value for data in highest class interval
	DMIN	Minimum value for data in lowest class interval
	DMP	Term used to locate interval which contains the mean
$\sqrt{D}$	DSQRT	The square root of D, see equation (2-15)
m	EM	A parameter in the Type IV Curve, see equation (2-30)
m_1	EM1	A parameter in Type I and VI Curves, see equation (2-23)
m_2	EM2	A parameter in Type I and VI Curves, see equation (2-23)
	EU	The independent normalized variable for the interval - 1/2 to + 1/2 based upon 16 subdivisions for Gaussian integration. The 16 values of EU are taken from Ref. 19 Chapter 2

Mathematical Notation	Fortran Notation	Description
	EZ	The term $(\sin\theta)^r e^{v\theta}$ used to determine F(R,V), see equation (2-33)
F		A ratio devised by Fisher to test for homogeneity of means, see equation (5-10)
fe		Degrees of freedom associated with Se^2, see equation (5-4)
	FRV	F(R, V) function, used for constant of Type IV curve, see Ref. 19, Chapter 2
G		A function used in the derivation of the constant from Type IV curve, see equation (2-33)
f	G(J)	Frequency of values in an interval
i	I	Index, usually denoting interval number
i		Term to indicate a complex root in derivation of Type IV curve, see equation (2-30)
	ID	Identification number for the field, well, and lease
	IDEPT	Depth of sample, feet
	INDEX	An index to distinguish permeability data cards from other data cards, INDEX = 1 for permeability INDEX = 2 for other data
	II	An index to distinguish the first cycle of a loop from all subsequent cycles
	IPERM	Permeability, millidarcys
ϕ	IPHI	Porosity, fractional
	ISAMP	Sample number
S_o	ISO	Oil saturation, fractional
∞		Infinity
	J	Index denoting number of different Pearson Type Curve fits desired

Mathematical Notation	Fortran Notation	Description
k		Number of variances considered in Bartlett's test
	L	Identification Program--numerical equivalent of AN in fixed point notation
	L	Fitting Program--index used in integration for YAVG
	MP	Number of the interval counting from the lowest which contains the mean
	NA	Number of intervals in fixed point notation
	NDEX	An index to distinguish among porosity, oil saturation and water saturation data cards: NDEX = 1 for porosity = 2 for oil saturation = 3 for water saturation
n_o		An estimate of an average sample size, see equation (5-8)
	NTYPE	Type of Pearson Curve to be calculated: NTYPE — Pearson Type 1 — I 2 — III 3 — IV 4 — VI
	NUMBR	Number of different type of curves to be calculated in a computer run
	P1	An intermediate variable for the Type IV curve
$\overline{P}$		A sample proportion, see equations (6-5) and (6-6)
q	PROP	Variable for an individual physical property such as porosity, permeability, etc.
π		An interval scale for the logrithms of permeability data
Q		A factor obtained from a table to compute D, see equation (5-12)

Mathematical Notation	Fortran Notation	Description
$\overline{q}$		Mean value of q, see equation (3-5)
Δq		The interval of q used to construct the histogram on the q scale
r	R	A parameter of the Pearson Type IV Curve, see equation (2-30)
r_1	R1	A parameter of the Pearson Type I and VI Curves, see equation (2-15)
r_2	R2	A parameter of the Pearson Type I and VI Curves, see equation (2-15)
S	S	A parameter of the Pearson Type IV Curve, see equation (2-30)
σ	SIGMA	The standard deviation of the set of data, see equation (3-9)
S^2		A term in Bartlett's test equal to $\dfrac{\Sigma\, v_i\, S_i^{\,2}}{v}$
$S_e^{\,2}$		Within sample or error variance, see equation (5-3)
$S_i^{\,2}$		Sample variance in Bartlett's test, see equations (5-13) and (5-14)
S^2 m, p		Variance of the population of means, see equation (5-5)
$S_p^{\,2}$		An estimate of the population variance, see equation (5-7) and (5-9)
$S_{\overline{p}}$		Standard deviation of the probability distribution of the sample proportion, see equation (6-5)
S_x		The sample standard deviation, see equation (6-2)
$S_x^{\,2}$		A sample variance, see equation (5-2)
$S_{\overline{x}}$		The standard deviation of the probability distribution of the sample means, see equation (5-11)

Mathematical Notation	Fortran Notation	Description
$\sigma_{\overline{x}}$		The standard deviation of the probability distribution of the sample mean, see equation (6-3)
	SUM	An intermediate storage cell used for the Gaussian integration in determining F(R, V)
	SUMD	Sum of data values in an interval
	SUMXG(I)	Sum for the I'th moment about the midpoint of the interval containing the mean, see Table IV
t	T	Independent variable for frequency distributions, in standardized notation, see equation (3-4)
	TANM1	Intermediate variable for computing YBAR of the Type Curve
	TEMP1	A temporary storage cell for an intermediate value in the calculation of the Type IV curve
	TEST	Term for determining whether a data value is within overall interval limits
θ	THETA	The angle equal to $\pi(EU) + \pi/2$ for use in the F(RV) calculation
	TOTAL	Total number of property samples for a field
	TR	Intermediate variable for computing YBAR of the Type Curve
μ_2	U2	2^{nd} moment about the mean, see equation (3-3)
μ_3	U3	3^{rd} moment about the mean, see equation (3-3)
μ_4	U4	4^{th} moment about the mean, see equation (3-3)
μ_r		The r^{th} moment about the mean, see equation (3-2)
μ_1'	UP1	First moment of the histogram about the arbitrary chosen midpoint, see equation (3-6)

Mathematical Notation	Fortran Notation	Description
μ_2'		Second moment of the histogram about the arbitrary chosen midpoint
μ_3'		Third moment of the histogram about the arbitrary chosen midpoint
μ_I'	UP(I)	I'th moment of the data about the midpoint of the interval which contains the mean, see equation (3-1)
$\overline{\mu}$		The mode of a given frequency distribution
X_2	UMOD2	2'nd moment about the mean incorporating Sheppard's corrections, see equation (3-10)
X_3	UMOD3	3'rd moment about the mean, incorporating Sheppard's corrections, see equation (3-11)
X_4	UMOD4	4'th moment about the mean, incorporating Sheppard's corrections, see equation (3-12)
v	V	A parameter of the Pearson Type IV Curve, see equation (2-30)
v		A term in Bartlett's test equal to $\Sigma\ v_i$
v_i		Degrees of freedom associated with each sample variance in Bartlett's test, see equation (5-13)
w		A parameter in the derivation of the constant for Type I Curve, see equation (2-26)
x_i	XBAR	The independent variable for the frequency distribution wherein each interval has unit width
	XMP	Equivalent of MP but in floating point notation
$\overline{X}_p$		A term used to compute $S^2_{m,\ p}$, see equation (5-6)
$\overline{X}_\phi$		The mean of a set of porosity data

Mathematical Notation	Fortran Notation	Description
$\overline{X}_{so}$		The mean of a set of oil saturation data
$\overline{X}_{sw}$		The mean of a set of water saturation data
	XX	The values of the independent variable for a frequency distribution corresponding to the left, middle, and right ends of an interval. These intervals are of unit width
	XZERO	The value of XX at the left extremity of a class interval
Y	Y	The frequency for a Pearson type curve at any value of the independent variable, t
	YAVG	Integrated average value of frequency over a specific interval, see equation (3-15)
	YNORM	A point value of frequency on the normal curve
Z	Z	A parameter for the Pearson Type VI curve, see equation (2-36)
z		The normal deviate corresponding to a desired confidence coefficient, see equation (6-6)
	ZERO	A numerical constant, equal to zero

TECHNICAL AND STATISTICAL TERMS

Term	Notation	Definition
Array		An arrangement of raw numerical data in ascending or descending order of magnitude
Chi-Square	χ^2	A measure of the discrepancy existing between observed and expected frequencies
Class Frequency		Number of individuals belonging to each class in summarizing large masses of raw data
Degrees of Freedom	d.f. fe v	The number N of independent observations in the sample (i.e., the sample size) minus the number K of population parameters which must be estimated from sample observations
Gamma Function	$\Gamma(X)$	A transcendental function used frequently in statistics and defined by equation (3-12)
Frequency Distribution		A tabular arrangement of data by classes together with the corresponding class frequencies
Histogram		A set of rectangles having: (a) bases on a horizontal axis (X axis) with centers at the class marks and lengths equal to the class interval sizes, (b) areas proportional to class frequencies
Mean	$\bar{X}$	Arithmetic average of a set of data
Median		The middle value of a set of numbers arranged in order of magnitude
Mode	$\bar{\mu}$	That value of a set of numbers which occurs with the greatest frequency
Moments	μ	A term used in the measurement of dispersion of a distribution and defined by equation (3-2)
Normal Distribution		Also referred to as Gaussian curve, normal curve of error, or normal probability distribution. A bell-shaped distribution defined by equation (2-40)

Term	Notation	Definition
Oil Saturation	S_o	A measure of the oil present within a rock expressed as a percentage of the total fluid saturation within a rock
Permeability	K	A property of a porous medium and a measure of the capacity of the medium to transmit fluids. Usually measured in darcies or millidarcies
Population		Term used in statistics to refer to the hypothetical complete enumeration of facts in a particular field of study
Porosity	ϕ	The ratio of the void space in a rock to the bulk volume of that rock multiplied by 100 to express it in percent
Range		The difference between the largest and smallest numbers in a set of numbers
Skewness		The degree of asymmetry, or departure from symmetry, of a distribution, see equation (2-13)
Standard Deviation	σ	The root mean square of the deviations from the mean
Standardized Unit	t	A term to express distance from mean to midpoint of histogram intervals in units of standard deviations, see equation (3-4)
Standardized Unit	z	A term to express the distance between the mean and another specified value of a frequency distribution as units of standard deviations
Unimodial		For frequency curves, those curves that have only one mode
Variance	σ^2	The square of the standard deviation
Water Saturation	S_w	A measure of the water present within a rock expressed as a percentage of the total fluid saturation within a rock

FORTRAN PROGRAMS AND FLOW CHARTS

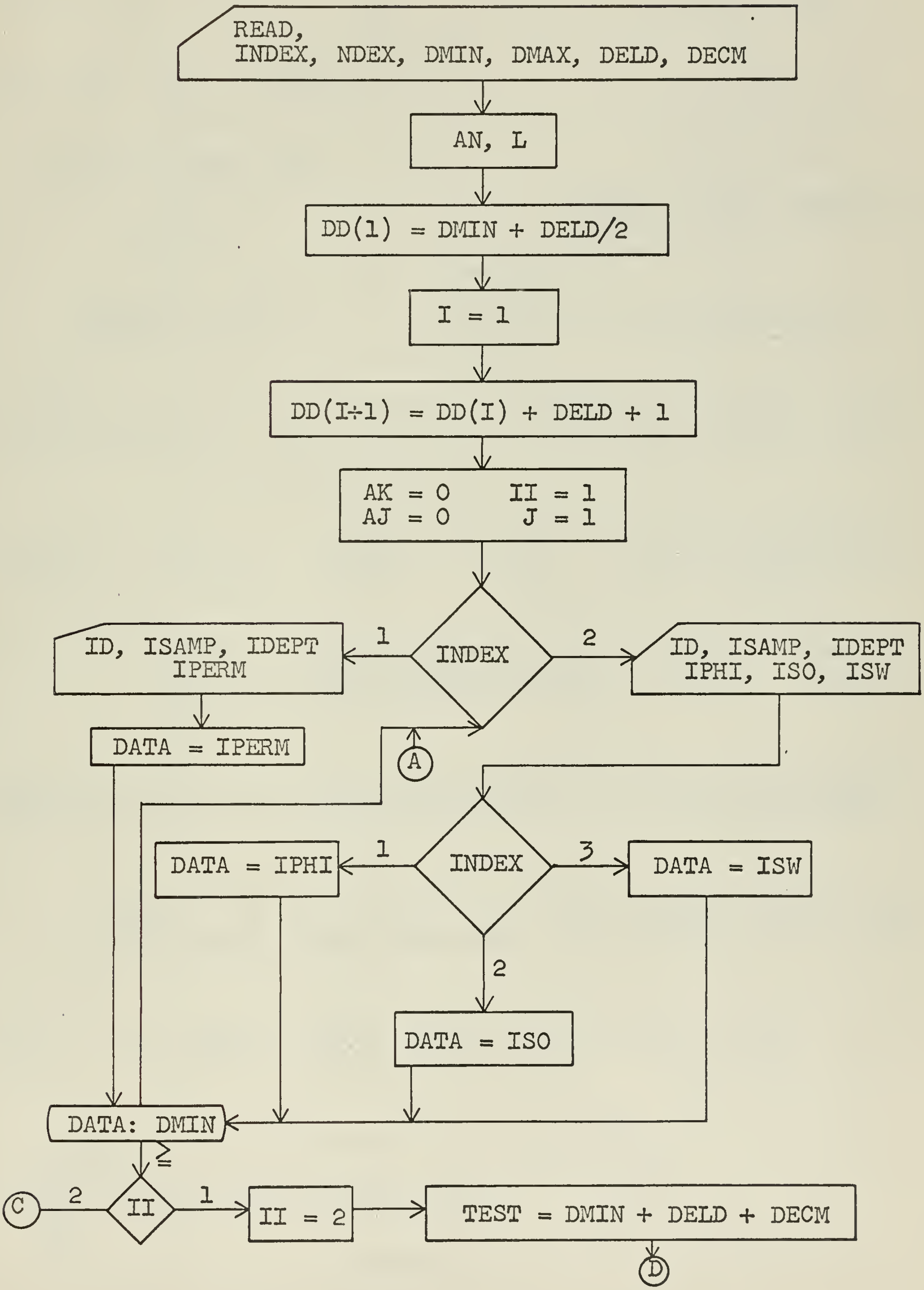
```
C     0000  0     650 FORTRAN PROGRAM TO
C     0000  0     CALCULATE A GOOD THEORETICAL             +
C     0000  0     FIT OF A PEARSON TYPE
C     0000  0     FREQUENCY CURVE FOR A GIVEN              -
C     0000  0     OBSERVED DISTRIBUTION.
C     0000  0     THE PROCEDURE TO BE FOLLOWED             +
C     0000  0      1 TABULATE THE DATA USING
C     0000  0        CONVENIENT CLASS INTERVALS.           +
C     0000  0      2 CALCULATE THE MOMENTS ABOUT           +
C     0000  0        A SELECTED VERTICAL.
C     0000  0      3 TRANSFER THE MOMENTS TO
C     0000  0         THE MEAN.
C     0000  0      4 APPLY SHEPPARDS CORRECTIONS           +
C     0000  0        TO THE MOMENTS.
C     0000  0      5 CALCULATE ALP3S ALP4 DELTA            -
C     0000  0      6 LOCATE THE MEAN (DAVG)
C     0000  0      7 DETERMINE FROM THE CHART
C     0000  0          WHAT TYPE OF CURVE TO USE
C     0000  0     PROGRAM BY J S VAN SCOYOC
C     0000  0     UNDER THE DIRECTION OF DR.
C     0000  0     FLOYD PRESTON

0     0000  0     DIMENSION DD(100),G(100),UP(8)           -
0     0000  1     ,SUMXG(8),X(100),XG(8)
0     0001  0     READ1, INDEX,NDEX
0     0030  0     READ1, DMIN, DMAX, DELD,DECM             +

C     0000  0     DETERMINE NUMBER OF INTERVALS            +

0     0040  0     AN =(DMAX-DMIN) /(DELD+DECM)             -
0     0041  0     IF(XCONF(1)) 42,50,50
0     0042  0     PUNCH1, AN
0     0050  0     L = XFIXF (AN)

C     0000  0     DETERMINE CENTER OF INTERVALS            +

0     0060  0     DD(1) = DMIN +DELD/2.0
0     0061  0     IF(XCONF(1))62,70,70
0     0062  0     PUNCH1, DD(1)
0     0070  0     DO 80  I = 1, L
0     0080  0     DD( I+ 1) =DD(I) + DELD+1.
0     0081  0     IF(XCONF(1)) 82,90,90
0     0082  0     PUNCH1, DD

C     0000  0     COUNT D,S IN EACH INTERVAL
C     0000  0     AND COUNT TOTAL D,S AND SUM D,           -

0     0090  0     AK=0.
0     0100  0     SUMD=0.
0     0121  0     II = 1
0     0122  0     J=1
0     0130  0     AJ=0.
0     0140  0     GO TO (150,180),INDEX

C     0000  0     READ IN DATA
```

```
0    0150  0     READ 1, ID, ISAMP, IDEPT,
0    0150  1       IPERM

C    0000  0    CONVERT TO FLOATING POINT

0    0160  0    DATA = FLOTF (IPERM)
0    0170  0    GO TO 270

C    0000  0    READ IN DATA

0    0180  0    READ1, ID, ISAMP, IDEPT, IPHI,          -
0    0180  1      ISO,ISW
0    0190  0    GO TO(200,220,240),NDEX

C    0000  0    CONVERT TO FLOATING POINT

0    0200  0    DATA = FLOTF (IPHI)
0    0210  0    GO TO 270
0    0220  0    DATA = FLOTF (ISO)
0    0230  0    GO TO 270
0    0240  0    DATA = FLOTF ( ISW)
0    0270  0    IF(DATA-DMIN)140,280,280
0    0280  0    GO TO(290,330),II
0    0290  0    II = 2
0    0300  0    TEST = DMIN + DELD+DECM
0    0330  0    IF (DATA-TEST) 340,380,380
0    0340  0    SUMD = SUMD + DATA
0    0360  0    AJ = AJ + 1.
0    0361  0    IF(XCONF(1))362,370,370
0    0362  0    PUNCH1,SUMD,DATA,TEST,AJ,AK,J,          -
0    0362  1    I
0    0370  0    GO TO 140
0    0380  0    AK = AK + AJ
0    0390  0    G(J) = AJ
0    0395  0    AJ=0.
0    0400  0    J = J + 1
0    0410  0    TEST = TEST + DELD+1.
0    0420  0    IF (TEST -(DMAX+DECM))330,330,
0    0420  1    430

C    0000  0    CALCULATE MEAN

0    0430  0    DAVG =  SUMD/ AK
0    0431  0    IF(XCONF(1))432,440,440
0    0432  0    PUNCH1,SUMD,AK,DAVG,G

C    0000  0    DETERMINE INTERVAL WHICH
C    0000  0    CONTAINS MEAN

0    0440  0    DMP = (DAVG-DMIN) /(DELD+DECM)          +
0    0450  0    MP = XFIXF (DMP) + 1
0    0460  0    XMP = FLOTF (MP)
0    0461  0    IF(XCONF(1))462,480,480
0    0462  0    PUNCH1, DMP,MP,XMP

C    0000  0    PLACE INTERVALS ON UNIT BASIS

0    0480  0    DO 490  I = 1, 8
0    0490  0    SUMXG(I) = 0.
```

```
0   0495  0   NA=L+1
0   0500  0   DO 560  I = 1,NA
0   0510  0   X(I) = FLOTF (I)-XMP
0   0511  0   IF(XCONF(1))512,520,520
0   0512  0   PUNCH1,X(I)

C   0000  0   CALCULATE XG TO X8G

0   0520  0   XG(1) = X(I) * G(I)
0   0530  0   DO 540 J = 1,7
0   0540  0   XG(J+1) = XG(J) * X(I)
0   0541  0   IF(XCONF(1))542,550,550
0   0542  0   PUNCH1, XG
0   0550  0   DO 560 J = 1, 8

C   0000  0   SUM XG TO X8G

0   0560  0   SUMXG(J) = SUMXG(J) + XG(J)                +
0   0561  0   IF(XCONF(1)) 562,570,570
0   0562  0   PUNCH1, SUMXG

C   0000  0   CALCULATE UPRIMES

0   0570  0   DO 580  I = 1, 8
0   0580  0   UP(I) = SUMXG (I) / AK
0   0581  0   IF(XCONF(1)) 582,590,590
0   0582  0   PUNCH1, UP

C   0000  0   CALCULATE U2 TO U8

0   0590  0   U2=UP(2)-UP(1)*UP(1)
0   0600  0   U3=UP(3)-3.*UP(1)*UP(2)+2.*                -
0   0600  1   UP(1)*UP(1)*UP(1)
0   0610  0   U4=UP(4)-4.*UP(1)*UP(3)+6.*                -
0   0610  1   UP(1)*UP(1)*UP(2)-3.*UP(1)*                -
0   0610  2   UP(1)*UP(1)*UP(1)
0   0620  0   SIGMA=U2**.5
0   0621  0   IF(XCONF(1)) 622,630,630
0   0622  0   PUNCH1,U1,U2,U3,U4

C   0000  0   APPLY SHEPPARDS CORRECTIONS

0   0630  0   UMOD2=U2-1./12.
0   0640  0   UMOD3=U3
0   0650  0   UMOD4=U4-0.5*U2+7./240.
0   0651  0   IF(XCONF(1)) 652,660,660
0   0652  0   PUNCH1,UMOD2,UMOD3,UMOD4

C   0000  0   CALCULATE VALUES NEEDED

0   0660  0   ALP3S=UMOD3*UMOD3/(UMOD2*UMOD2             -
0   0660  1   *UMOD2)
0   0670  0   ALP3=ALP3S**.5
0   0680  0   ALP4=UMOD4/(UMOD2*UMOD2)
0   0690  0   DELTA=(2.*ALP4-3.*ALP3S-6.)/               +
0   0690  1   (ALP4+3.)
0   0760  0   DELD=DELD+DECM
0   0770  0   PUNCH1,NA,N,DMIN,DMAX,DELD,
0   0770  1   DAVG
```

```
0     0780  0     DO790 I=1,NA
0     0785  0     PUNCH1,I,DD(I),X(I),G(I)
0     0790  0     CONTINUE
0     0795  0     PUNCH1,ALP3,ALP3S,DELTA,SIGMA,                 +
0     0795  1     UP(1)
0     0810  0     GO TO 30
0     0820  0       END
```

FLOW CHART TO IDENTIFY THE PEARSON TYPE CURVE CORRESPONDING TO A GIVEN DATA SET



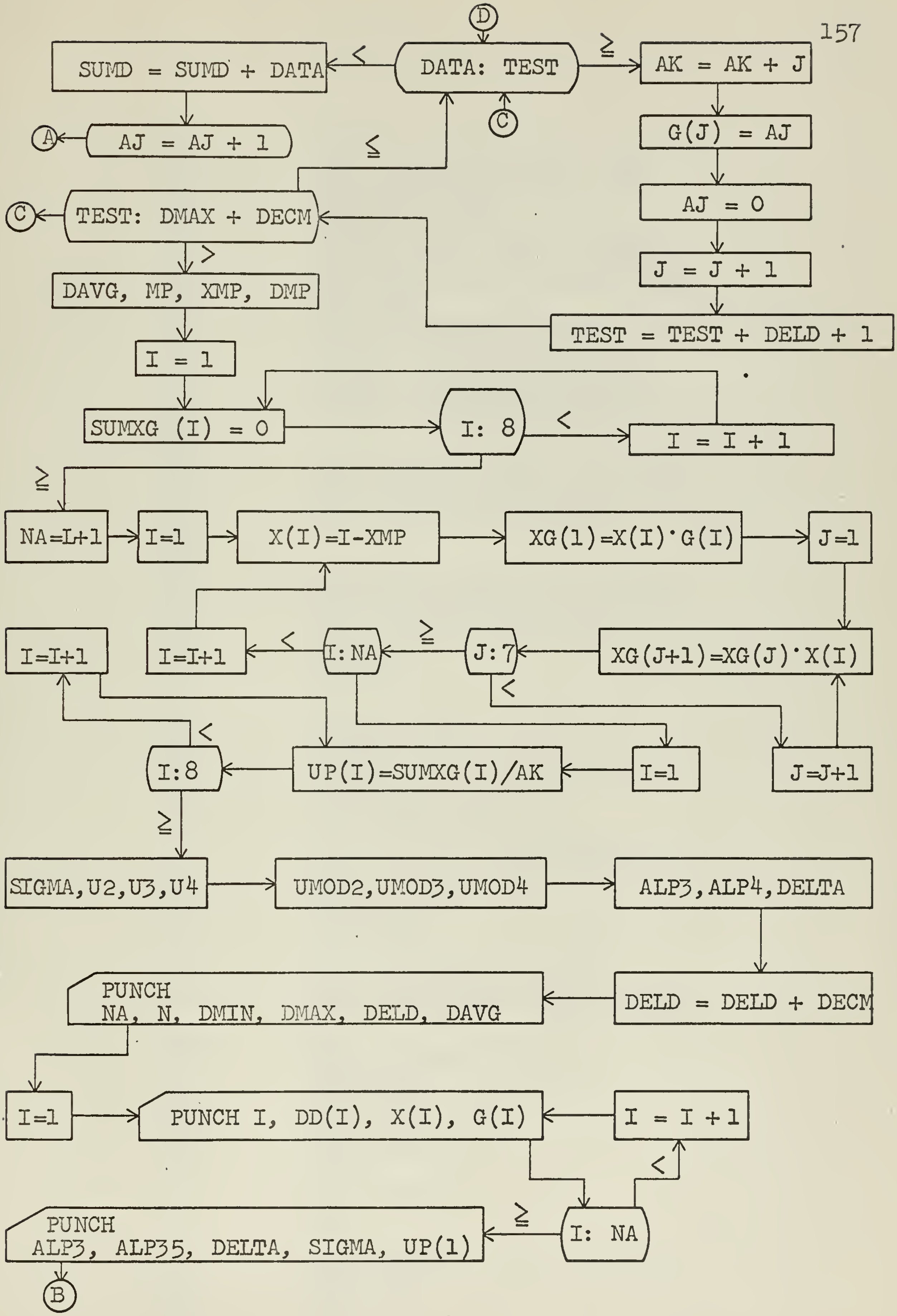

D
DATA: TEST
SUMD = SUMD + DATA
AK = AK + J
AJ = AJ + 1
A
C
G(J) = AJ
AJ = 0
TEST: DMAX + DECM
J = J + 1
DAVG, MP, XMP, DMP
TEST = TEST + DELD + 1
I = 1
SUMXG (I) = 0
I: 8
I = I + 1
NA=L+1
I=1
X(I)=I-XMP
XG(1)=X(I)·G(I)
J=1
I=I+1
I=I+1
I: NA
J: 7
XG(J+1)=XG(J)·X(I)
I:8
UP(I)=SUMXG(I)/AK
I=1
J=J+1
SIGMA,U2,U3,U4
UMOD2,UMOD3,UMOD4
ALP3,ALP4,DELTA
PUNCH
NA, N, DMIN, DMAX, DELD, DAVG
DELD = DELD + DECM
I=1
PUNCH I, DD(I), X(I), G(I)
I = I + 1
PUNCH
ALP3, ALP35, DELTA, SIGMA, UP(1)
I: NA
B

FIGURE 31

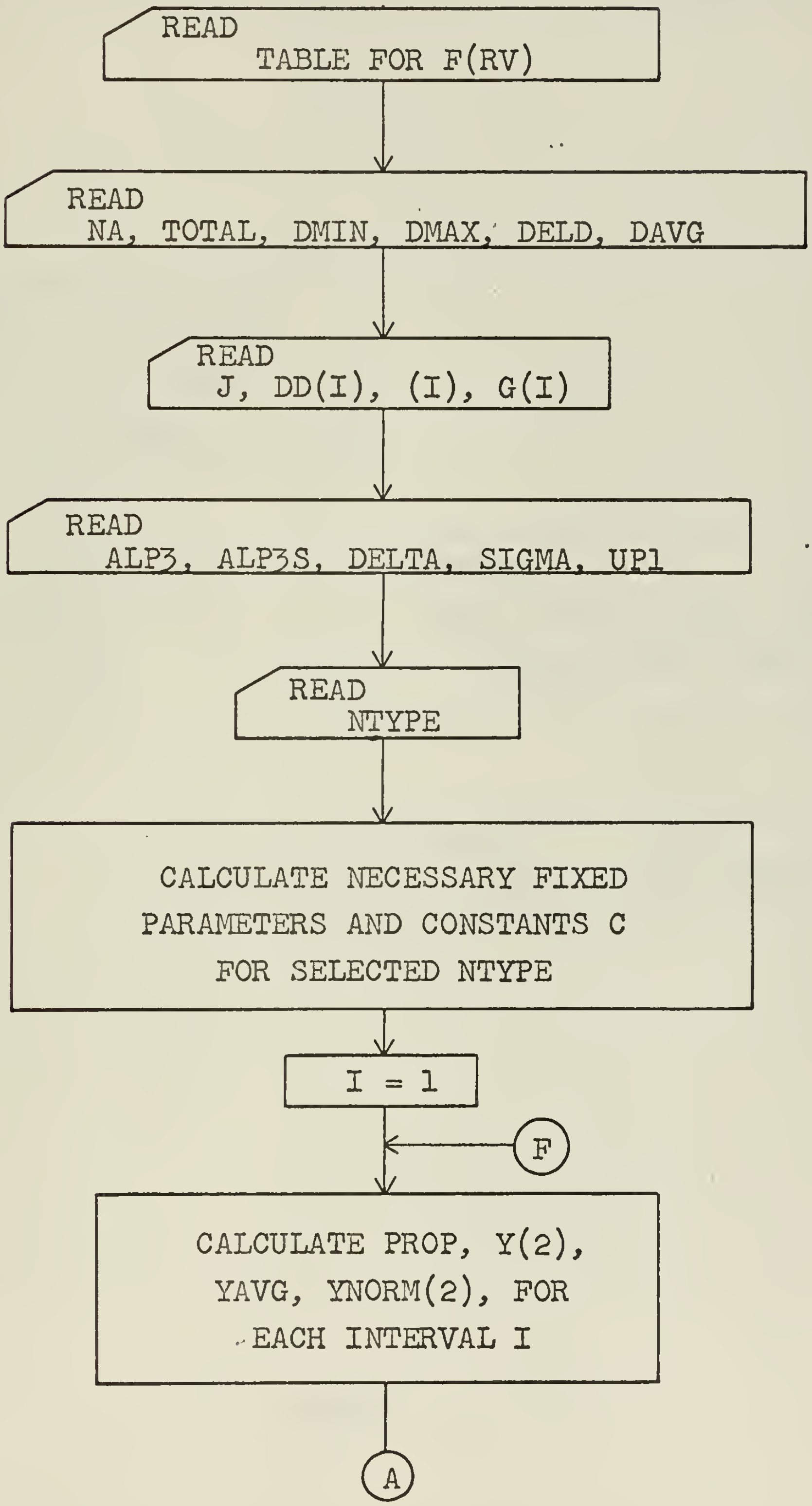
```
C     0000  0    PEARSON FIT FOR TYPES 1,3,4,6
C     0000  0    DATA FOR READ2,READ3, AND
C     0000  0    READ4, ARE SUPPLIED BY THE
C     0000  0    IDENTIFICATION
C     0000  0    NUMBER (FOR READ5,)=NUMBER
C     0000  0    OF DIFFERENT TYPES OF CURVES          +
C     0000  0    WHICH ARE TO BE FITTED TO THE
C     0000  0    SAME DATA
C     0000  0    NTYPE=CURVE TYPE NUMBER
C     0000  0    NTYPE=1 FOR PEARSON TYPE 1
C     0000  0    NTYPE=2 FOR PEARSON TYPE 3
C     0000  0    NTYPE=3 FOR PEARSON TYPE 4
C     0000  0    NTYPE=4 FOR PEARSON TYPE 6
C     0000  0
C     0000  0    THIS PROGRAM NEEDS SINF AND           +
C     0000  0    XCONF SUBROUTINES AND SPECIAL         +
C     0000  0    GAMMA FUNCTION ROUTINE
C     0000  0
C     0000  0
0     0000  0    DIMENSION DD(50),X(50),G(50),
0     0000  1    XX(3),Y(3),YNORM(3),CON(5),
0     0000  2    CST(16),EU(16),EZ(16),
0     0000  3    THETA(16)
0     0001  0    ZERO=0.

0     0000  0    READ1,CST,EU
0     0000  0    DO 501 I=1,16
0     0000  0    THETA(I)=3.1415927*EU(I)+
0     0000  1    1.5707963
0     0501  0    EZ(I)=SINF(THETA(I))

0     1011  0    READ2,NA,TOTAL,DMIN,DMAX,DELD,        +
0     1011  1    DAVG
0     0000  0    AN=FLOTF(NA)
0     0003  0    DO4I=1,NA

0     0004  0    READ3,J,DD(I),X(I),G(I)
0     0005  0    READ4,ALP3,ALP3S,DELTA,SIGMA,         -

0     0005  1    UP1
0     0006  0    D=ALP3*ALP3-4.*DELTA*
0     0006  1    (DELTA+2.)
0     0007  0    IF(D)8,9,9
0     0008  0    D=-D
0     0009  0    DSQRT=D**.5

0     0010  0    READ5,NUMBR
0     0011  0    J=1

0     0012  0    READ6,NTYPE
0     0013  0    DO14I=1,5
0     0014  0    CON(I)=0.
0     0015  0    PUNCH1,NTYPE
```

```
0    0000  0    NPCH=417
0    0000  0    IF(XCONF(1))417,17,17
0    0417  0    PUNCH1,NPCH,NTYPE,NUMBR,NA,D,
0    0417  1    DSQRT,AN
0    0017  0    GO TO(18,30,36,18),NTYPE

0    0018  0    EM1=-((1.+DELTA)/DELTA)*(1.-                +
0    0018  1    ALP3/DSQRT)-1.
0    0019  0    CON(1)=EM1
0    0020  0    EM2=-((1.+DELTA)/DELTA)*(1.+                +
0    0020  1    ALP3/DSQRT)-1.
0    0021  0    CON(2)=EM2
0    0022  0    R1=(-ALP3+DSQRT)/(2.*DELTA)                 +
0    0023  0    CON(3)=R1
0    0024  0    R2=(-ALP3-DSQRT)/(2.*DELTA)                 +
0    0025  0    CON(4)=R2
0    0026  0    GO TO(27,29,29,48),NTYPE
0    0027  0    C=TOTAL/(SIGMA*(R2-R1)**(EM1+               -
0    0027  1    EM2+1.)*GAMAF(EM1+1.)*GAMAF(EM              +
0    0027  2    2+1.))
0    0127  0    CON(5)=C
0    0028  0    GO TO 51
0    0029  0    STOP

0    0030  0    A=2./ALP3
0    0131  0    CON(1)=A
0    0031  0    ASQ=A*A
0    0032  0    C=A**(ASQ)/(SIGMA*EXPEF(ASQ)*               -
0    0032  1    GAMAF(ASQ))
0    0034  0    CON(2)=C
0    0035  0    GO TO 51

0    0036  0    R=ALP3/(2.*DELTA)
0    0037  0    CON(1)=R
0    0038  0    EM=1./DELTA+2.
0    0039  0    CON(2)=EM
0    0040  0    S=DSQRT/(2.*DELTA)
0    0041  0    CON(3)=S
0    0042  0    P1=2.*EM-2.
0    0043  0    V=-2.*(1.+DELTA)*ALP3/(DELTA*
0    0043  1    DSQRT)
0    0044  0    CON(4)=V
0    0146  0    TEMP1=2.*EM-1.
0    0000  0    SUM=0.
0    0000  0    DO 502 I=1,16
0    0000  0    Y=(EZ(I)**R)*EXPEF(V*THETA(I))
0    0000  0    SUM=SUM+Y*CST(I)
0    0502  0    FRV=EXPEF(-1.5707963*V)*SUM*                -
0    0502  1    3.1415927
0    0046  0    C=TOTAL*S**TEMP1/(SIGMA*FRV)
0    0246  0    CON(5)=C
0    0047  0    GO TO 51

0    0048  0    ALPHA=R1-R2
0    0049  0    C=TOTAL*GAMAF(-EM2)/(SIGMA*                 -
0    0049  1    GAMAF(EM1+1.)*GAMAF(-EM2-EM1-               -
0    0049  2    1.)*ALPHA**(EM1+EM2+1.))
0    0050  0    CON(5)=C
```

```
0    0051  0    CHISQ=0.
0    0000  0    NPCH=452
0    0000  0    IF(XCONF(1))452,52,52
0    0452  0    PUNCH1,NPCH,CON
0    0052  0    BELSQ=0.
0    0053  0    XZERO=X(1)-.5
0    0153  0    PROP=DMIN+DELD*.5
0    0054  0    I=1
0    0055  0    XX(1)=XZERO
0    0056  0    XX(2)=X(I)
0    0057  0    XX(3)=XZERO+1.
0    0058  0    L=1
0    0059  0    XBAR=XX(L)
0    0060  0    T=(XBAR-UP1)/SIGMA
0    0061  0    YNORM(L)=.3989421*EXPEF(-T*T/2              -
0    0061  1    .)*TOTAL/SIGMA
0    0062  0    GO TO(63,68,70,65),NTYPE

0    0063  0    YBAR=(C*(T-R1)**EM1)*((R2-T)**
0    0063  1    EM2)
0    0064  0    GO TO 91

0    0065  0    Z=(T-R2)
0    0066  0    YBAR=C*Z**EM2*(Z-ALPHA)**EM1                -
0    0067  0    GO TO 91

0    0068  0    YBAR=C*(A+T)**(ASQ-1.)*EXPEF                +
0    0068  1    (-A*T)
0    0069  0    GO TO 91

0    0070  0    TR=(T+R)
0    0071  0    ARG=TR/S
0    0072  0    XSQ=ARG*ARG
0    0000  0    COE=3.
0    0073  0    IF(XSQ-1.)83,175,76
0    0175  0    TANM1=.78539816
0    0275  0    GO TO 89
0    0076  0    TANM1=1.5707963-1./ARG
0    0077  0    Q=1.
0    0078  0    DO81I=1,9
0    0079  0    TANM1=TANM1+Q/(ARG*XSQ*COE)                 +
0    0080  0    XSQ=XSQ*ARG*ARG
0    0000  0    COE=COE+2.
0    0081  0    Q=-Q
0    0082  0    GO TO 89
0    0083  0    TANM1=ARG
0    0084  0    Q=-1.
0    0085  0    DO 88I=1,9
0    0086  0    TANM1=TANM1+Q*ARG*XSQ/COE
0    0087  0    XSQ=XSQ*ARG*ARG
0    0000  0    COE=COE+2.
0    0088  0    Q=-Q
0    0089  0    ARG1=V*(1.5707963-TANM1)
0    0090  0    YBAR=C*((TR*TR+S*S)**(-EM)*                 -
0    0090  1    EXPEF(ARG1))

0    0091  0    Y(L)=YBAR
0    0000  0    NPCH=492
0    0000  0    IF(XCONF(1))492,92,92
```

```
0     0492  0     PUNCH1,NPCH,I,L,XX,YBAR,T,
0     0492  1     YNORM,XSQ,TANM1
0     0092  0     IF(L-3)93,95,95
0     0093  0     L=L+1
0     0094  0     GO TO 59

0     0095  0     YNAVG=(YNORM(1)+YNORM(3)+4.*                 +
0     0095  1     YNORM(2))/3.
0     0095  1     YNORM(2))/6.
0     0096  0     YAVG=(Y(1)+Y(3)+4.*Y(2))/6.                  +
0     0097  0     PUNCH1,I,X(I),PROP,G(I),Y(2),
0     0097  1     YAVG,YNORM(2)
0     0098  0     IF(I-NA)99,111,111

0     0099  0     BSQ=G(I)-YNAVG
0     0100  0     BSQ=BSQ*BSQ
0     0000  0     IF(YNAVG) 101,102,101
0     0101  0     BELSQ=BELSQ+BSQ/YNAVG
0     0102  0     CHI=G(I)-YAVG
0     0103  0     CHI=CHI*CHI
0     0000  0     IF(YAVG) 104,105,104
0     0104  0     CHISQ=CHISQ+CHI/YAVG
0     0105  0     XZERO=XZERO+1.
0     0106  0     Y(1)=Y(3)
0     0107  0     YNORM(1)=YNORM(3)
0     1107  0     PROP=PROP+DELD
0     0000  0     NPCH=4109
0     0000  0     IF(XCONF(1))4109,109,109
0     4109  0     PUNCH1,NPCH,I,L,CHI,CHISQ,BSQ,               +
0     4109  1     BELSQ
0     0109  0     I=I+1
0     0110  0     GO TO 55

0     0111  0     PUNCH1,ZERO
0     0112  0     PUNCH1,DMIN,DMAX,DELD,DAVG,
0     0112  1     SIGMA,ALP3,DELTA
0     0113  0     PUNCH1,ZERO
0     0114  0     PUNCH1,CON,CHISQ,BELSQ
0     0115  0     IF(J-NUMBR)116,121,121
0     0116  0     J=J+1
0     0118  0     PAUSE

0     0119  0     GO TO 12
0     0121  0     PAUSE
0     0122  0     GO TO 1011
0     0123  0     END
```

GENERALIZED FLOW CHART FOR CURVE FIT TO
PEARSON TYPE DISTRIBUTION



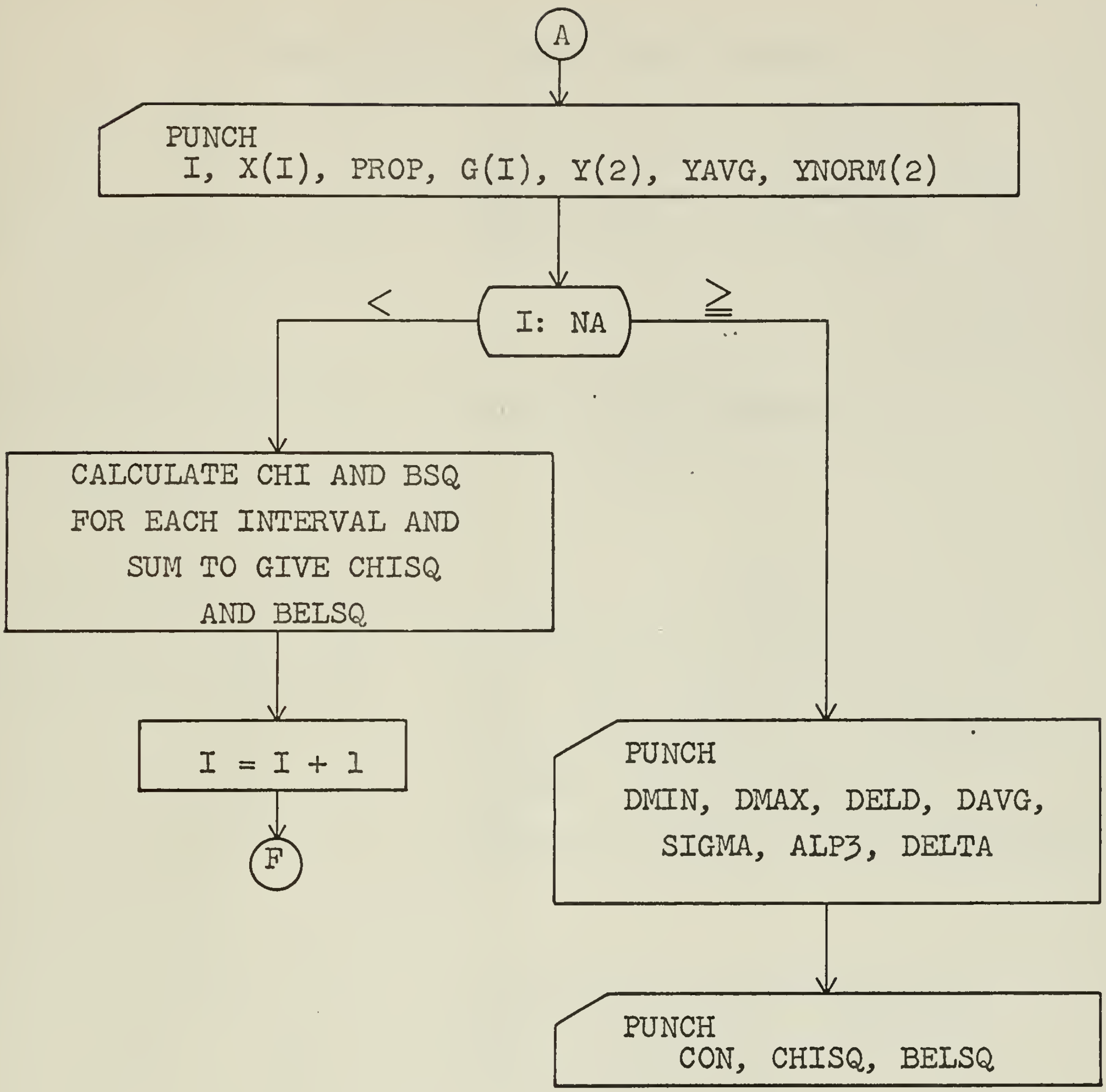
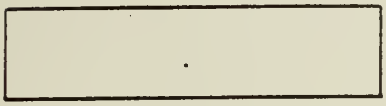

A
PUNCH
I, X(I), PROP, G(I), Y(2), YAVG, YNORM(2)
I: NA
<
≧
CALCULATE CHI AND BSQ
FOR EACH INTERVAL AND
SUM TO GIVE CHISQ
AND BELSQ
I = I + 1
F
PUNCH
DMIN, DMAX, DELD, DAVG,
SIGMA, ALP3, DELTA
PUNCH
CON, CHISQ, BELSQ

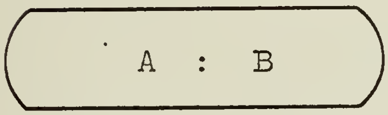
FIGURE 32

EXPLANATION OF FLOW CHART SYMBOLS

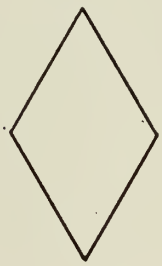
A Read or Punch instruction. Data are read or punched in the order given, left to right.



An arithmetic calculation.



A test of the difference A - B for negativity, positivity, or zero. The program branches to the next appropriate calculation sequence depending upon the numerical value of this difference.



A multi-way branch depending upon the numerical value of the inclosed index.

A jump is made in the flow chart to the circle with the same enclosed letter.

FIGURE 33

APPENDIX C

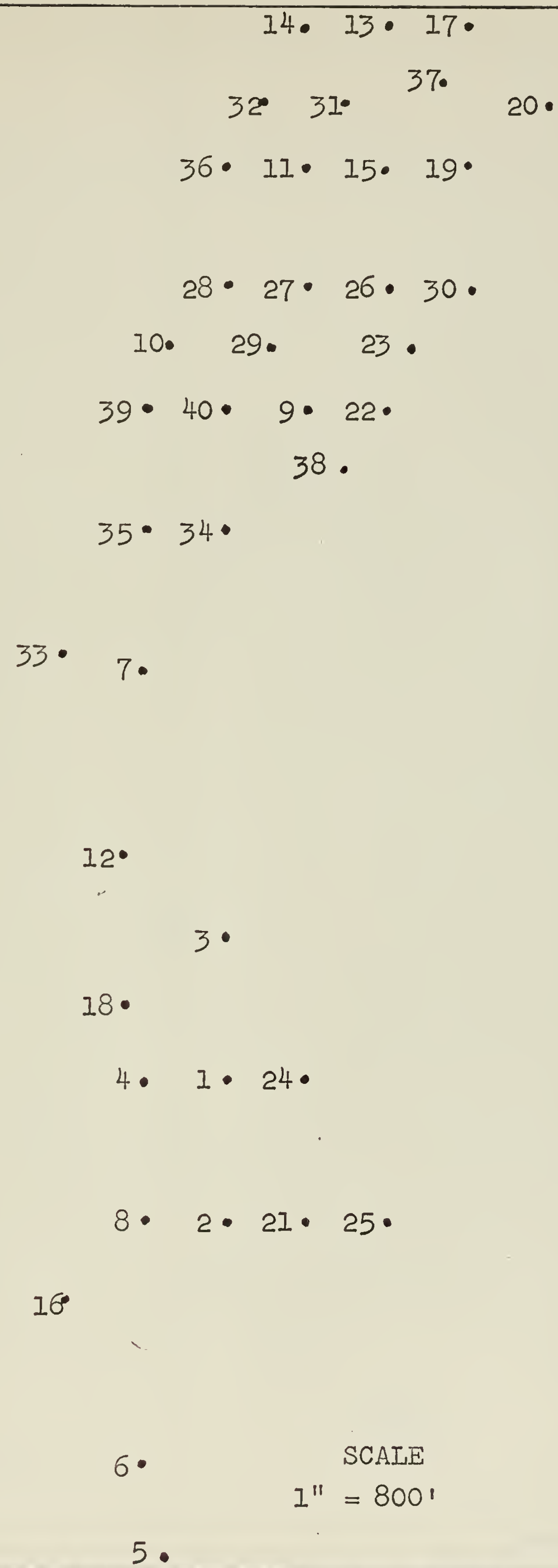


FIGURE 34 MAP OF WELLS CORED--FIELD 1

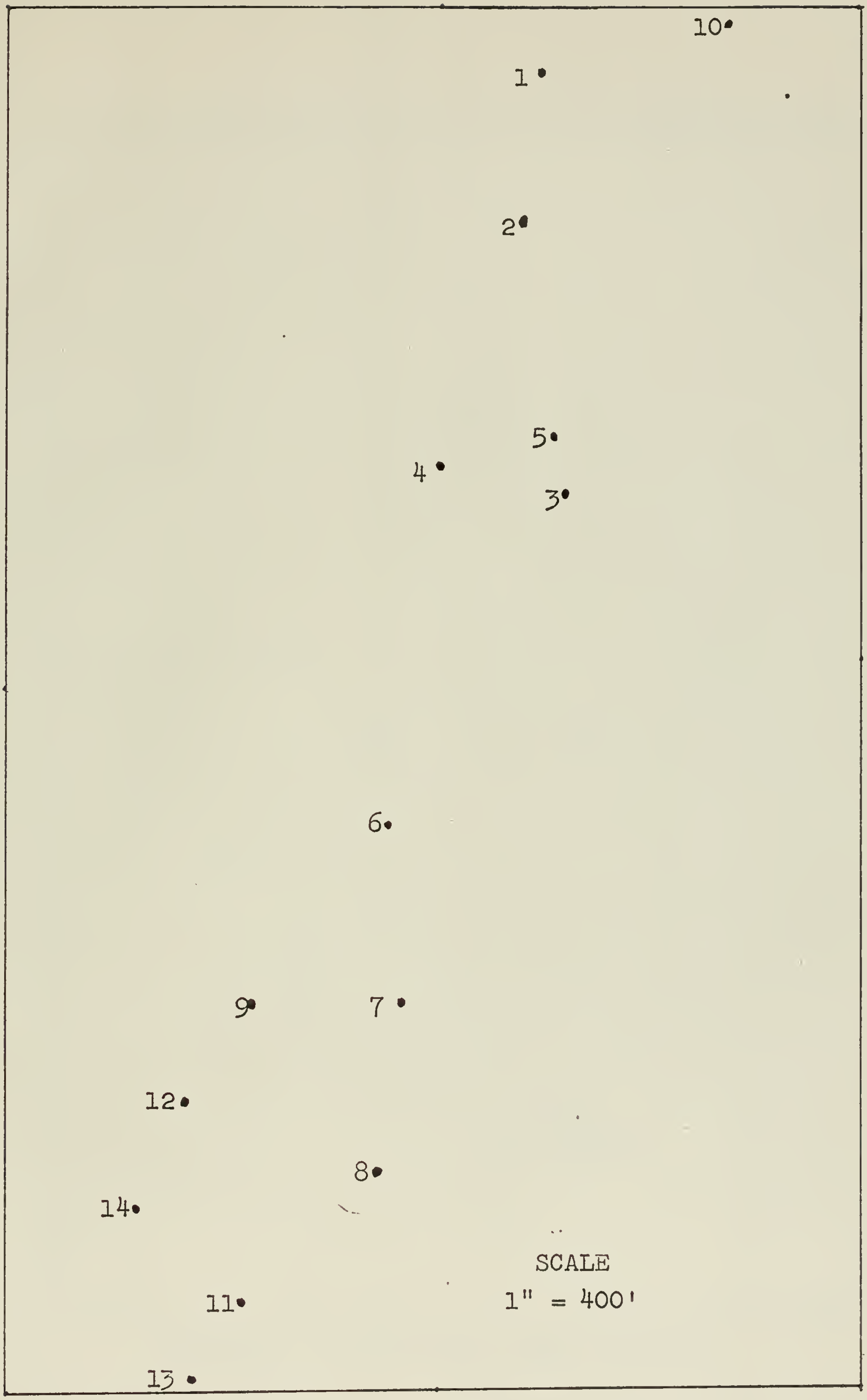


FIGURE 35 MAP OF WELLS CORED--FIELD 2

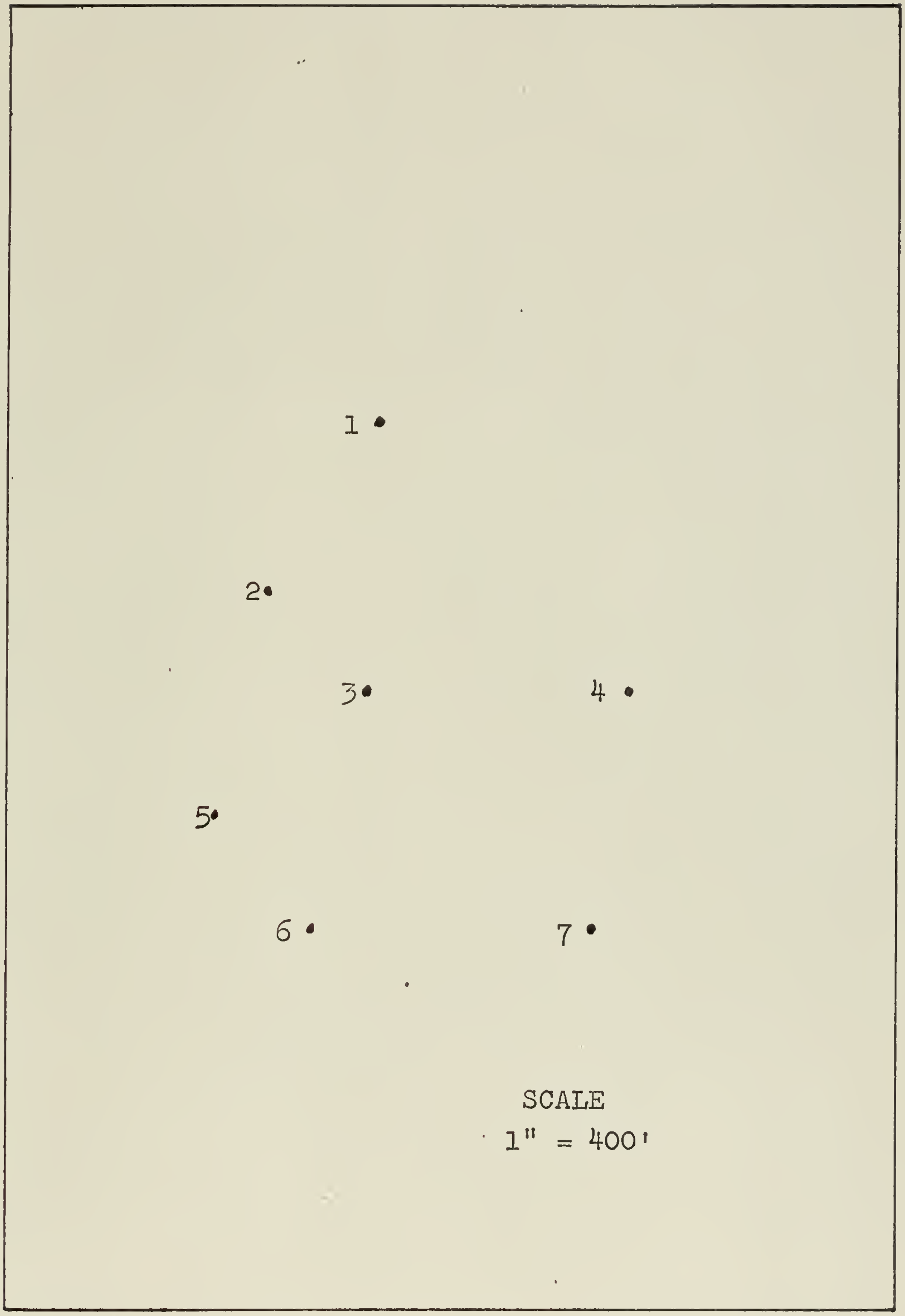


FIGURE 36 MAP OF WELLS CORED--FIELD 3